Etude de l'influence de la qualité audiovisuelle sur la qualité d'expérience du spectateur : combinaison d'indicateurs subjectifs, physiologiques et oculaires

Julie Lassalle

N° d'ordre : 2013telb0299

Sous le sceau de l'Université européenne de Bretagne

Télécom Bretagne

En habilitation conjointe avec l'Université de Bretagne-Sud

Ecole Doctorale – SICMA

ETUDE DE L'INFLUENCE DE LA QUALITE AUDIOVISUELLE SUR LA QUALITE D'EXPERIENCE DU SPECTATEUR : Combinaison d'indicateurs subjectifs, physiologiques et oculaires

---

Thèse de Doctorat

Mention : Sciences et Techniques de l'Information, de la Communication et de la Connaissance

Présentée par **Julie Lassalle**

Département : LUSSI
Laboratoire : Lab-STICC
Pôle : CID

Directeur de thèse : Gilles Coppin

Soutenue le 22 octobre 2013

Jury :

M. Patrick Le Callet, Professeur, Ecole polytechnique, Nantes (Rapporteur)
M. Charles Tijus, Professeur, Université Paris 8 (Rapporteur)
M. Eric Jamet, Professeur, Université Rennes 2 (Examinateur)
Mme Janick Naveteur, Maitre de conférences, Université de Valenciennes (Examinateur)
M. Gilles Coppin, Professeur, Télécom Bretagne, Brest (Directeur de thèse)
Mme Laetitia Gros, Docteur, Orange Labs, Lannion (Encadrante)
M. Thierry Morineau, Professeur, Université Bretagne-Sud, Vannes (Co-directeur de thèse)

À André et Louise
À ma famille
À Guillaume

REMERCIEMENTS

Cette thèse a été réalisée grâce au soutien et à la collaboration de nombreuses personnes, sans pouvoir les citer toutes, je leur adresse mon entière reconnaissance pour leur contribution.

En premier lieu, je remercie Patrick Le Callet et Charles Tijus pour avoir rapporté ce travail ainsi que Janick Naveteur et Eric Jamet pour avoir accepté d'en être les examinateurs.

Je tiens ensuite à remercier chaleureusement Laëtitia Gros pour son encadrement rigoureux, son soutien et son implication tout au long de ces années et Gilles Coppin pour avoir été un directeur de thèse dont l'énergie, la sagacité et le soutien n'ont jamais fait défaut. Je remercie également Thierry Morineau, co-directeur de ce travail, pour avoir accepté d'attraper au vol ce projet ainsi que pour ses conseils avisés et indispensables à l'élaboration de ce document. Je mesure la chance d'avoir bénéficié d'un encadrement de qualité tant sur le plan humain que professionnel et qui a toujours su fédérer mes idées et orienter la thèse dans le bon sens.

Je remercie bien évidement Orange Labs et particulièrement Gwenaël Le Lay pour m'avoir accueilli au sein de son équipe et pour avoir hébergé, dans les meilleures conditions, ce projet de thèse. Je ne saurai oublier l'aide de Jean-Charles Guicquel, Emmanuel Wyckens, Bernard Letertre, Jean-Yves Leseure, Catherine Quinquis, Julien Libouban ou encore Mickael Bonin pour le temps passé à m'épauler sur les aspects techniques incombant à la mise en place du matériel expérimental audiovisuel et aux traitements des séquences de test.

L'analyse des mesures psychophysiologiques a demandé la prise en main d'environnements de « programmation » inconnus pour moi auparavant. Je ne remercierais jamais assez Romain Deprez pour la patience et la pédagogie dont il a fait preuve pour m'initier à ces nouveaux instruments tout en sacrifiant une partie de son temps de doctorant. Cette thèse a en effet évolué dans un climat d'entraide et de solidarité entre doctorants et je tiens à remercier plus particulièrement Yves et Maty. Je n'aurai pu espérer meilleurs collègues de bureau.

Mes remerciements sincères vont également à Jérôme Daniel, Ronan Lepage, Jean-Marc Goujon, Françoise Fessant, Sorin Moga et Philippe Lenca pour leur travail et le temps accordé au traitement des données.

Ce travail de thèse n'aurait pu aboutir sans l'aide précieuse et le soutien technique de Jean-Marc Diverrez que je remercie amicalement pour les heures passées à résoudre les problèmes logiciels et matériels des expérimentations et pour avoir toujours répondu à mes sollicitations.

Enfin, je veux dédier en priorité ce travail à ma famille et à mon conjoint pour avoir vécu cette thèse en même temps que moi. Merci de tout cœur pour votre présence et votre soutien indéfectible.

Selon Isaac Asimov « *la Science-Fiction est la branche de la littérature qui se soucie des réponses de l'être humain au progrès de la science et de la technologie* », là est aussi l'objectif de certaines branches de la psychologie …

RESUME

Etude de l'influence de la qualité audiovisuelle sur la qualité d'expérience du spectateur : combinaison d'indicateurs subjectifs, physiologiques et oculaires.

Dans un contexte fortement concurrentiel, l'un des principaux enjeux pour les acteurs de l'offre de services audiovisuels (AV) est de garantir au spectateur une *qualité d'expérience* (QoE) optimale. Aujourd'hui, la QoE est souvent restreinte à la perception de la qualité audiovisuelle restituée (QAV) par le système. Elle est principalement mesurée à travers la collecte de notes données par des participants sur des échelles de qualité, après visualisation et écoute de séquences AV traitées par le ou les technologies à évaluer. Ces tests subjectifs suivent des procédures recommandées par l'Union Internationale des Télécommunications. Cependant, la qualité restituée peut affecter la QoE (fatigue, effort, *etc.*) sans être reflétée par les notes de qualité. Une méthode considérant l'évaluation non plus de la QAV perçue seule mais de la *qualité d'expérience*, plus largement considérée, pourrait permettre de mieux rendre compte de l'influence de la qualité du son et de l'image sur le spectateur. Le présent travail est centré sur la recherche d'une méthode alternative aux méthodes actuelles de l'évaluation de qualité pour applications multimédias dans un contexte de visualisation et d'écoute de contenus AV 2D ou 3D. L'approche proposée aborde la QoE sous l'angle de l'analyse conjointe d'indicateurs subjectifs et d'indicateurs toniques physiologiques (activité électrodermale, rythme cardiaque, température cutanée périphérique, volume sanguin périphérique) et oculaires (PERCLOS, durée et fréquence de clignement de l'œil, nombre de saccades, diamètre pupillaire). Les mesures physiologiques et oculaires ont pour avantage de ne pas être assujetties aux biais des mesures subjectives (représentativité, échelles, *etc.*) et de traduire des phénomènes comme la fatigue ou l'effort mental, potentiellement induits par la présence de dégradations sur les signaux audio et/ou vidéo et pouvant être critiques du point de vue de la QoE. Deux protocoles ont été testés. Les résultats ont montré que la QAV module l'expérience subjective et que les notes de qualité ne sont pas suffisantes pour refléter fidèlement cet effet. L'influence de la QAV sur les mesures physiologiques et oculaires est moins évidente. Un ensemble de facteurs, notamment lié à certains attributs du contenu de test comme la dynamique ou la luminosité, aurait pu masquer ou atténuer l'observation d'un effet de la qualité restituée. Néanmoins, deux des indicateurs physiologiques ont réagi à la présence de dégradations audio et/ou vidéo lorsque celles-ci étaient cumulées à l'effet préjudiciable d'autres facteurs (vidéo 3D ou effet de passation).

Mots-clefs : Qualité d'expérience, qualité audiovisuelle, évaluation, mesures subjectives, mesures physiologiques, mesures oculaires, fatigue mentale, effort mental.

SUMMARY

Study of the influence of the audiovisual quality on the spectator quality of experience: combination of subjective, physiological and ocular indicators.

In a strongly competitive context, one of the main stakes for the actors of the audiovisual services offer is to guarantee to the spectator an optimal Quality of Experience (QoE). Nowadays, QoE is often limited to the perception of the audiovisual quality (AVQ) received by the system. It is mainly measured through the collection of rates given by testers onto quality scales, after visualization and listening of the AV sequences treated by one or several technologies to be evaluated. These subjective tests are following procedures recommended by the International Telecommunication Union. However, the restored quality can affect some factors of QoE (fatigue, effort, *etc*.) which are not reflected by the quality scores. A method considering the evaluation either of the AV quality only received but of the quality of experience, widely considered, could allow to report better the influence of the sound and image quality on the spectator. The present work is centered on the research of an alternative method to current methods of quality assessment for multimedia applications in a context of viewing/listening of 2D or 3D AV contents. The proposed approach addresses the QoE in terms of analysis of subjective indicators and tonic physiological (electrodermal activity, heart rate, peripheral cutaneous temperature, blood volume pulse) and oculars indicators (PERCLOS, duration and frequency of the blinking of the eye, number of saccadics movements, pupillary diameter). Physiological and ocular measures have for advantage not to be subjected to the biases of the subjective measures (representativeness, scales, *etc*.) and to reflect phenomena such as fatigue or mental effort, potentially induced by the presence of audio and/or video degradations, wich may be critical in terms of QoE. Two protocols were tested to study the relevance of this approach. The results showed that QAV modulates the subjective measures and have putted forward the insufficiency of quality rates to reflect faithfully this effect. The impact of the quality on the physiological and ocular measures is less obvious. A set of factors in particular connected to certain attributes of the test contents, as the dynamics or the luminosity, would have been able to mask or decrease the quality effects observation on gathered measures. However, two of the physiological indicators reacted to the presence of audio and/or video degradations when these were accumulated to the detrimental effect of other factors (3D video or test period effect).

Keywords: Audiovisual quality, quality of experience, subjective measures, physiological measures, ocular measures, mental fatigue, mental effort.

LISTE DES ABREVIATIONS

AED : Activité ElectroDermale

AV : AudioVisuel

DP : Diamètre Pupillaire

EBdur : Eye blink duration

Ebfreq : Eye Blink Frequence

FC : Fréquence Cardiaque

QAV : Qualité AudioVisuelle

QoE : Quality of Experience

SAC : Saccade

SNA : Système Nerveux Autonome

SNP : Système Nerveux Parasympathique

SNS : Système Nerveux Sympathique

TCP : Température Cutanée Périphérique

UIT : Union Internationale des Télécommunications

VRC : Variation du Rythme Cardiaque

VSP : Volume Sanguin Périphérique

TABLES DES MATIERES

TABLE DES FIGURES

LISTE DES TABLEAUX

INTRODUCTION

Dans un contexte extrêmement concurrentiel, la *qualité d'expérience* de l'utilisateur (QoE : Quality of Experience) est une des préoccupations principales des acteurs du domaine de l'offre de services audiovisuels (télévisuels, visioconférences, *etc.*). Actuellement, l'évaluation de la QoE se réalise généralement à travers l'évaluation de la qualité, telle que perçue par les utilisateurs, des signaux audio et/ou vidéo restitués. Les méthodes d'évaluation utilisées sont recommandées par l'Union Internationale des Télécommunications. Ces approches reposent sur des mesures subjectives dont l'interprétation et la validité sont limitées par un certain nombre de biais. Elles ne permettent pas non plus de rendre compte fidèlement de l'influence de la qualité audiovisuelle sur la *qualité d'expérience* de l'utilisateur. Par exemple, ces méthodes n'apportent pas d'informations sur le coût pour l'utilisateur, du point de vue de la fatigue ou de l'effort mental, induit par des dégradations du signal et pouvant à terme conduire à un rejet du système ou de la technologie de restitution. Le *coût utilisateur* peut être mesuré à partir d'indices de l'activité physiologique et oculaire de l'individu. Ce type de mesures présente l'avantage de ne pas être soumis aux biais des mesures subjectives, capables de diminuer la fiabilité des réponses recueillies (problème de représentativité, interprétation des items, références implicites, *etc.*). Wilson G. M. et Sasse (2000a, 2000b) ont montré que des fluctuations importantes de qualité audio ou vidéo peuvent ne pas être consciemment perçues par les participants et pour autant être reflétées par l'activité physiologique. L'objectif du travail présenté dans ce document est de contribuer à une méthode alternative pour l'évaluation de la qualité audiovisuelle pour applications multimédias. La démarche a consisté à proposer, au sein d'une même méthode, l'étude conjointe à la fois de mesures subjectives, avec un questionnaire enrichi (c.-à-d. évaluation de critères autres que les seules notes de qualité) et de mesures de l'activité physiologique et oculaire du spectateur. Cette approche doit permettre de mieux comprendre l'influence de la qualité audiovisuelle restituée sur la qualité d'expérience du spectateur, d'une part, en palliant les problèmes des méthodes actuelles d'évaluation et d'autre part, en proposant une approche plus holistique de l'influence de la qualité sur la QoE.

Le **premier chapitre** présente le cheminement du signal audiovisuel de sa captation à sa restitution. Il définit ensuite la notion de qualité du signal audiovisuel restitué par les systèmes de transmission, tout d'abord, sous l'angle de ses caractéristiques physiques puis sous l'angle de la perception du spectateur. Enfin, les différentes méthodes standardisées et recommandées par l'Union Internationale des Télécommunications pour l'évaluation de la qualité audiovisuelle seront présentées.

Le **second chapitre** est consacré à la perception bimodale audiovisuelle. L'influence de la qualité audiovisuelle sur la perception du spectateur relève des capacités des systèmes auditifs et visuels humain à traiter et interpréter l'information bimodale. La perception de la qualité audiovisuelle repose sur la manière dont les stimuli auditifs et visuels sont fusionnés pour former un percept unique de qualité globale audiovisuelle. Les disparités temporelles et spatiales capables d'interférer voire d'empêcher le processus d'intégration audiovisuel (et à terme d'entraîner une diminution de la qualité perçue) seront abordées. Une dernière partie

sera dédiée à l'étude des influences mutuelles entre qualité audio et vidéo ainsi que leur contribution respective à la qualité audiovisuelle globale perçue par le spectateur.

Le **chapitre 3** décrit le fonctionnement général du système nerveux humain. Les indicateurs physiologiques et oculaires étudiés dans ce document ainsi que leurs mesures et traitements statistiques seront ensuite détaillés.

Le **chapitre 4** expose les différentes influences capables de moduler l'activité physiologique et/ou oculaire d'un individu. Les influences émotionnelles, attentionnelles puis celles liées à l'effort mental et à la fatigue seront abordées.

Le **chapitre 5** montre l'intérêt des mesures psychophysiologiques (physiologiques et oculaires) pour évaluer le coût, pour l'individu, de l'utilisation d'un service ou d'une technologie. Plus spécifiquement, le lien entre mesures psychophysiologiques et évaluation de qualité sera abordé. Les études intégrant ces mesures aux méthodes d'évaluation de la qualité seront utilisées dans ce document comme référent méthodologique. Enfin, une méthode hybride incluant à la fois la mesure de l'expérience subjective, de l'activité tonique physiologique et oculaire du spectateur sera proposée comme alternative aux méthodes actuelles d'évaluation de la qualité audiovisuelle.

Le **chapitre 6** est dédié à la présentation d'une expérimentation exploratoire testant une première proposition de méthode hybride. Des contenus de test 2D présentant un niveau fluctuant de qualité étaient présentés à un panel de participants. Il était attendu que les dégradations influencent autant les mesures subjectives que les mesures psychophysiologiques (du point de vue de l'effort mental et de la fatigue). L'analyse des données psychophysiologiques, essentiellement à partir d'analyses de variance, portait sur l'activité tonique de quatre indicateurs physiologiques et quatre indicateurs oculaires. Les résultats ont mis en avant un effet de la qualité sur les mesures subjectives. La qualité a influencé un seul des indicateurs physiologiques et oculaires (activité électrodermale) et ce, pour un seul des contenus présentés. Cette première étude a mis l'accent sur la sensibilité des mesures psychophysiologiques à différents facteurs susceptibles de masquer ou d'atténuer les effets de qualité. Différentes propositions ont été émises à la suite de ces résultats pour réadapter le protocole proposé. Une attention particulière doit notamment être portée au contenu qui a influencé tant les mesures subjectives que psychophysiologiques.

Le **chapitre 7** est consacré à la caractérisation des contenus de test. Dans un premier temps, un expert audiovisuel a réalisé une annotation précise, à partir d'un ensemble des descripteurs, des contenus du corpus de test. Ensuite, un échantillon de séquences extraites des contenus caractérisés a été proposé à un panel de participants naïfs. Cette étape a permis de vérifier, à partir de l'étude de la concordance entre caractérisation experte et naïve, la pertinence d'un sous-ensemble de descripteurs considérés comme pouvant varier selon la perception de l'individu qui l'évalue. Des critères propres à l'expérience du spectateur (intérêt, plaisir, *etc*.) ont aussi été évalués afin de mieux comprendre l'influence du contenu sur la *qualité d'expérience*. Enfin, ces mêmes séquences ont été présentées, dans une seconde

expérimentation, avec des dégradions de qualité dans l'intention d'étudier le lien entre contenu et perception de la qualité audiovisuelle. Ces expériences ont notamment mis en avant une influence des descripteurs « expression sonore », « modalité » et « dynamique » sur la perception de qualité. Ces informations sont utiles pour optimiser la sélection des séquences de test et pour améliorer la compréhension de l'influence de la qualité sur les mesures subjectives et psychophysiologiques.

Enfin, le **chapitre 8** propose un protocole tenant compte des conclusions de la première expérimentation (chapitre 6), c'est-à-dire des facteurs considérés comme potentiellement préjudiciables à l'expression physiologique et/ou oculaire des dégradations. Un questionnaire enrichi est également proposé pour obtenir un retour plus complet de l'influence de dégradations sur la qualité perçue et la *qualité d'expérience* du spectateur. Ce questionnaire a en partie été élaboré à partir des expérimentations présentées dans le chapitre 7. Des contenus de test 3D présentant un niveau fluctuant de qualité étaient présentés à un panel de participants. Il était attendu que les dégradations influencent autant les mesures subjectives que les mesures psychophysiologiques (effort mental, fatigue). L'analyse des données psychophysiologiques portait sur l'activité tonique de quatre indices physiologiques et cinq indices oculaires essentiellement à partir d'analyses de variance. Les résultats de cette expérimentation ont mis en avant un effet de la qualité sur les mesures subjectives. La qualité a influencé un seul des indicateurs physiologiques et oculaires (fréquence cardiaque) et ce, pour un seul des contenus présentés. Cette expérimentation a également montré que les notes de qualité reportées selon les recommandations de l'UIT ne permettent pas de rendre compte de l'ensemble de l'influence de la qualité sur la *qualité d'expérience* du spectateur. Ce dernier résultat confirme l'importance de faire évoluer les méthodes actuelles vers des méthodes reflétant plus fidèlement l'influence de la qualité audiovisuelle sur la *qualité d'expérience*.

Finalement, cette thèse repose sur l'étude d'indicateurs subjectifs et sur la capacité des mesures psychophysiologiques à évaluer le coût (effort/fatigue) pour le spectateur induit par la présence de dégradations sur les signaux audio et/ou vidéo restitués avec comme perspective la proposition d'une méthode hybride capable de refléter plus fidèlement, que les méthodes actuelles, l'influence de la qualité audiovisuelle sur la *qualité d'expérience* du spectateur.

CHAPITRE I – QUALITE AUDIOVISUELLE ET EVALUATION

1.1. ENJEUX

Il y a environ quatre-vingts ans, la diffusion en noir et blanc de contenus audiovisuels apparaissait, suivie quarante ans plus tard par la couleur. Depuis, le degré de technicité n'a cessé de croître. Aujourd'hui, des systèmes de restitution audiovisuelle de haute qualité (format haute définition -HD- ou 3D -images et sons-) s'étendent à l'ensemble des terminaux (mobiles, ordinateur, télévision) et services (communications interpersonnelles, télévisuelles -TV-, *etc*.) disponibles. L'évolution des technologies de diffusion audiovisuelle tend à proposer des contextes de visualisation et d'écoute toujours plus immersifs où les sens du spectateur sont stimulés de façon nouvelle pour s'approcher de la réalité.

Face à cette constante évolution et dans un contexte fortement concurrentiel, un des principaux enjeux pour les acteurs de l'offre de services audiovisuels (AV) est de garantir une *qualité d'expérience* (QoE -Quality of Experience-) optimale à l'utilisateur. Cela est particulièrement vrai pour les nouvelles technologies à forte valeur ajoutée, telles que l'audio et la vidéo HD ou 3D. Ainsi, une attention toute particulière est portée à l'évaluation de la *qualité d'expérience* à travers le développement d'outils et la mise en place de méthodes d'évaluation. Pour les fournisseurs de services audiovisuels, la QoE est notamment étudiée à travers la perception de la qualité des médias (c.-à-d. la qualité des signaux audio et/ou vidéo restitués à l'utilisateur). La qualité perçue, et plus largement la QoE, devient un élément clé qu'il faut par conséquent savoir mesurer.

Actuellement, les méthodes utilisées pour mesurer la qualité perçue par l'utilisateur sont généralement celles recommandés par l'Union Internationale des Télécommunications (UIT) dont l'objectif est de proposer un cadre commun d'évaluation aux différents laboratoires, instituts ou entreprises cherchant à évaluer la qualité perçue de service de restitution audio et/ou vidéo (téléphonie, TV, visioconférence, *etc*.). Ces méthodes reposent principalement sur la collecte de notes subjectives recueillies sur des échelles de qualité, après visualisation et/ou écoute de séquences traitées par le service ou la technologie à évaluer.

Toutefois, ces méthodes comportent certains biais inhérents aux mesures subjectives. Par exemple, la tâche même d'évaluation de qualité par les participants est peu représentative d'une situation réelle de visualisation. De la même manière, les échelles et les catégories utilisées pour juger la qualité peuvent biaiser la réponse des participants. Ces biais tendent à fragiliser la validité et l'interprétation des mesures recueillies. Les mesures subjectives limitent également l'étude de l'influence de la qualité restituée sur la QoE. Par exemple, le coût pour l'utilisateur, en matière d'effort ou de fatigue, lors de la visualisation et de l'écoute de contenus audiovisuels dont le signal audio et/ou vidéo est dégradé n'est pas reflété par les méthodes subjectives proposées par l'UIT.

Les méthodes normalisées ne permettent donc pas d'obtenir un retour suffisamment représentatif de l'influence des systèmes ou technologies AV sur l'utilisateur. Les services (TV, visioconférence, mobile, *etc.*) existants ou le déploiement de nouveaux services doivent tenir compte de l'influence de la qualité sur la perception de l'utilisateur mais aussi sur la qualité de son expérience afin de proposer le service le plus adapté possible. Ainsi, les méthodes actuelles d'évaluation de la qualité audiovisuelle doivent être adaptées et élargies à la QoE pour répondre aux besoins d'évaluation des technologies et services innovants. Différentes études (Durin, Gros et Chateau, 2006 ; Gros, Chateau, Durin, 2006 ; Gros, Chateau et Macé, 2005 ; Wilson G. M. et Sasse, 2000a, 2000b) proposent une approche alternative basée sur l'évaluation non plus à travers le jugement seul de qualité effectué de façon explicite par les participants mais à travers l'impact de la qualité des médias sur le comportement, l'état émotionnel ou l'activité physiologique des utilisateurs. Une méthode alternative pourrait donc consister à ajouter ce type de mesures aux méthodes subjectives actuelles d'évaluation de qualité.

L'objectif de ce chapitre est d'aborder la notion de qualité audiovisuelle, restituée par les systèmes de transmission, et telle que perçue par l'utilisateur. Les avantages et les faiblesses des méthodes subjectives actuelles pour évaluer la qualité seront présentés.

1.2. TRANSMISSION DU SIGNAL

Le niveau de qualité audiovisuelle restituée est conditionné par l'ensemble des traitements appliqués au signal audiovisuel, de la production à la restitution, en passant par la transmission.

Généralement, les techniques de production audiovisuelle permettent d'obtenir des contenus de très bonne qualité (qualité source). Cependant, les conditions et techniques de prise de vues peuvent dégrader ce niveau de qualité. Par exemple, dans le cas de la 3D, les systèmes de caméra ne sont pas toujours adaptés et une mauvaise utilisation peut générer des artefacts gênants pour la vision humaine. De la même manière, certains contenus 2D sont convertis en 3D grâce à des outils de post-production automatiques qui utilisent des modèles de reconstruction 3D pouvant être mal adaptés et aboutir à un rendu moins confortable pour le spectateur.

La phase de transmission est beaucoup plus critique du point de vue de la qualité restituée. Qu'elle soit hertzienne (TNT, satellitaire, *etc*.) ou filaire (*via* internet, fibre optique, *etc.*), elle permet la diffusion de contenus audiovisuels à partir d'un émetteur (par exemple une tête de réseau) vers différents terminaux (TV, mobile, ordinateur). La quantité de données du signal source étant trop grande par rapport aux capacités des canaux de transmission, des algorithmes de compression sont donc appliqués au signal numérique avant l'émission. Ces traitements génèrent des pertes d'informations définitives plus ou moins importantes par rapport au signal source, impactant ainsi la qualité perçue du signal audiovisuel par l'utilisateur. Au niveau de la réception, le signal est décodé et restitué sur l'écran et les haut-

parleurs du terminal. Les outils de décodage sont en général bien maîtrisés et délivrent une bonne qualité. Cependant, la partie restitution ou adaptation du signal décodé peut générer des artefacts liés à la colorimétrie, au sous-échantillonnage, au dé-entrelacement, autant d'éléments qui peuvent dégrader la qualité perçue.

La norme de compression actuellement utilisée pour diffuser de la TVHD ou du contenu multimédia multi-supports est la norme MPEG4-AVC, également connue sous le sigle H264. Le débit moyen de diffusion est variable selon l'application visée et le mode de transmission. Actuellement, en TVHD, la vidéo est codée aux alentours de 6 Mégabits par seconde (Mbps) pour l'ADSL et son débit moyen pour la TNT est d'environ 11 Mbps. Le signal audio est lui codé entre 128 Kilobits par seconde (Kbps) et 384 Kbps selon le mode du signal (mono, stéréo ou multicanal) et la norme de codage utilisée (adaptée au mode). A l'émission, le signal codé (ou compressé) est partitionné en paquets de données contenant du signal audio et/ou vidéo ainsi que des données additionnelles pouvant être utilisées au niveau du décodeur pour gérer la qualité du signal, alimenter des applications d'enrichissement de la TV, *etc*. Au cours du transport, des paquets de données peuvent être corrompus ou perdus donnant lieu à différents types d'artefacts, selon les mécanismes de gestion de paquets perdus : coupures, « voix de robot » sur le signal audio, apparition de blocs de pixels sur le signal vidéo, *etc*.

Chaque étape du traitement de la chaîne audiovisuelle, de la réalisation du contenu à sa restitution, peut introduire une perte d'information ou une distorsion altérant la qualité du signal restitué.

1.3. QUALITE DU SIGNAL

La qualité objective du signal peut être définie par des caractéristiques techniques mesurables. Plus précisément, ce sont les performances techniques du système de transmission qui vont être mesurées. Les performances obtenues vont permettre de déterminer le niveau de la qualité du signal et *in fine* de service (QoS -Quality of Service-). La QoS désigne la capacité d'un système à fournir un service conforme aux exigences de diffusion audiovisuelle en matière de délai ou de débit à fournir par exemple.

Cette approche dite *centrée technologie* repose sur un ensemble de critères de qualité dont les principaux sont : les paquets d'information audio et/ou vidéo (paquets d'information manquants), le débit (volume d'informations maximal -bits- par unité de temps), la bande-passante audio (c.-à-d. de la bande de fréquence -KHz- utilisée pour transmettre le signal audio, conditionne la fidélité de la restitution), la gigue (fluctuation du signal dans le temps ou en phase) ou encore le délai/latence (temps mis par un paquet d'information entre l'émission et la réception). L'ensemble de ces paramètres va influencer le niveau de qualité du signal restitué.

Des performances faibles de l'un ou de plusieurs de ces paramètres (réduction bande-passante, réduction du débit, perte de paquets, *etc.*) diminuent la qualité de service. Les dégradations engendrées par la diminution des performances, liées aux conditions de transmission, vont altérer la qualité du signal restitué en bout de chaîne et par conséquent, conduire à des dégradations audio et/ou vidéo potentiellement perceptibles par l'utilisateur.

Un des axes majeurs de recherche des instituts de télécommunications consiste à évaluer dans quelle mesure les dégradations du signal diminuent le niveau de qualité audiovisuelle perçue par l'utilisateur. La Figure 1.1 ci-dessous illustre le cheminement du signal audiovisuel de sa production à la qualité finale perçue.

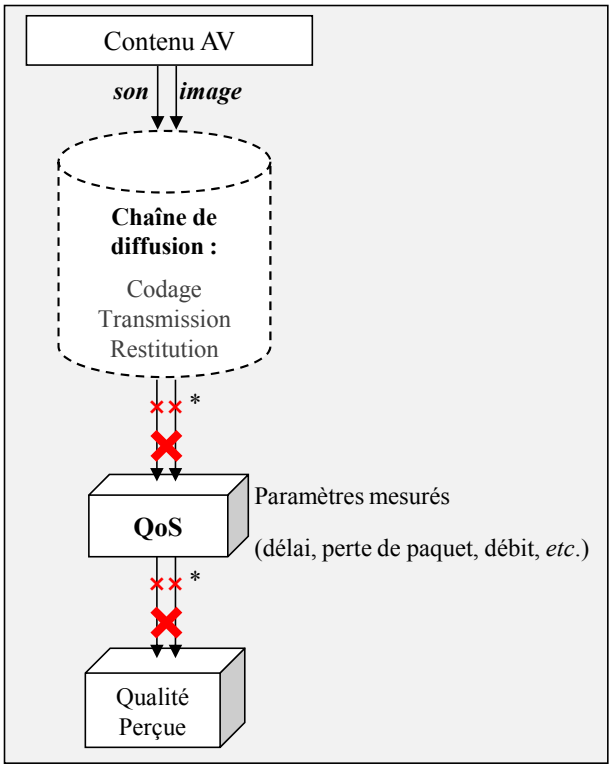


Fig. 1.1 Cheminement du signal audiovisuel de sa production à la qualité finale perçue par l'utilisateur.

La présence de dégradations peut conduire à une altération de la qualité du signal, ainsi restitué, et être à l'origine d'une diminution du niveau de qualité perçue par l'utilisateur.

1.4. DE LA QUALITE PERÇUE A LA QUALITE D'EXPERIENCE

Les paramètres de QoS ne permettent pas toujours de caractériser l'influence des dégradations liées à la baisse de leurs performances sur la perception de l'utilisateur final.

Le terme de qualité, considéré du point de vue de sa perception, qu'il soit associé à l'image ou au son (parole, musique ou bandes sonores) est utilisé pour désigner des notions différentes, bien que liées.

Le terme de qualité peut être défini comme la **caractérisation perceptive intrinsèque** d'un signal. Par exemple, la sonie (intensité perçue du signal sonore), la hauteur (fréquence fondamentale d'un son), le timbre (perception distincte entre deux sons d'intensité et de

hauteur équivalentes), la dynamique (écart entre un son fort et un son faible) ou la localisation spatiale des sources permettent de caractériser l'audio. Pour la vidéo, des attributs de contraste, de définition des contours, de résolution et de température de couleur (proportion de rouge, de vert et de bleu d'une image : une forte proportion de bleu est qualifiée de couleur *froide* tandis qu'une lumière *chaude* présente une proportion élevée de rouge) peuvent être considérés comme des dimensions de la qualité d'une image.

Mais le terme de qualité peut également caractériser la qualité de transmission d'un système qui capture, traite, transmet et/ou reproduit les signaux. Dans ce cas, les qualificatifs souvent utilisés sont ceux qui caractérisent les **altérations du signal original** par le système, tels que distordu, bruité, flou, saccadé, *etc*.

Selon l'application (c.-à-d. le système évalué), l'une et/ou l'autre de ces deux définitions de la qualité peut être envisagée : celle du rendu intrinsèque sans référence à l'original et celle de la fidélité en référence au signal original.

Le monde des télécommunications s'intéresse davantage à l'impact de l'ensemble de la chaîne de transmission sur le signal original, c'est-à-dire de l'émission du signal (en passant par les phases de codage/décodage et le canal de transmission) au terminal de restitution. La qualité du signal est alors considérée du point de vue de l'utilisateur des services de restitution audiovisuelle.

Pastrana-Vidal (2005) suggère de réunir les différents attributs perceptifs caractérisant les dégradations du signal (saccades, flou, *etc.*) autour de deux grands axes de perception de qualité : la netteté (c.-à-d. notion de contraste et de définition des contours) et la fluidité pour la vidéo, la clarté et la fluidité pour l'audio. La présence de dégradations audio ou vidéo affecterait ces différents axes dont l'altération diminuerait le niveau de qualité perçue par l'utilisateur.

Par exemple, la continuité (silences, « clics », *etc.*) ou la clarté (distorsions, échos) du signal audio peut être altérée lorsque le débit est trop faible ou en présence de pertes de paquets d'informations. Du point de vue de la vidéo, des pertes de paquets ou la réduction du débit peuvent aussi être à l'origine d'une altération du flux vidéo en matière de fluidité (saccades, gel de l'image) et par conséquent de la qualité du signal restitué. De la même manière, la netteté de la vidéo pourrait être altérée par la présence de blocs (groupe de pixels) ou de flou (réduction de la netteté des contours et des détails spatiaux par dégradation globale de l'image) sur certaines zones de l'image, en raison, par exemple, d'un débit insuffisant. A titre d'illustration, la Figure 1.2 ci-dessous présente une image (extraite d'un signal vidéo) dégradée par la présence de blocs de pixels.

Fig. 1.2. Illustration d'une image dégradée (image de droite) par la présence de blocs suite à une diminution du débit vidéo. L'image originale est présentée à gauche.

En plus de ces dégradations unimodales, le signal audiovisuel, c'est-à-dire le signal global considéré comme un évènement unique, peut également être altéré par des dégradations lui étant spécifiques. Deux types de dégradations AV peuvent survenir entre les modalités audio et vidéo : les dégradations AV spatiales et les dégradations AV temporelles. Les dernières, généralement provoquées par un délai (latence) générée par le processus de transmission du signal, correspondent à la désynchronisation image/son et sont fréquemment rencontrées dans les contextes actuels de visualisation de contenu audiovisuel. Les dégradations AV spatiales sont moins liées à des erreurs de transmission du signal qu'à un écart angulaire trop important entre les dispositifs de restitution audio et vidéo. Ces disparités seront plus amplement décrites dans le chapitre II.

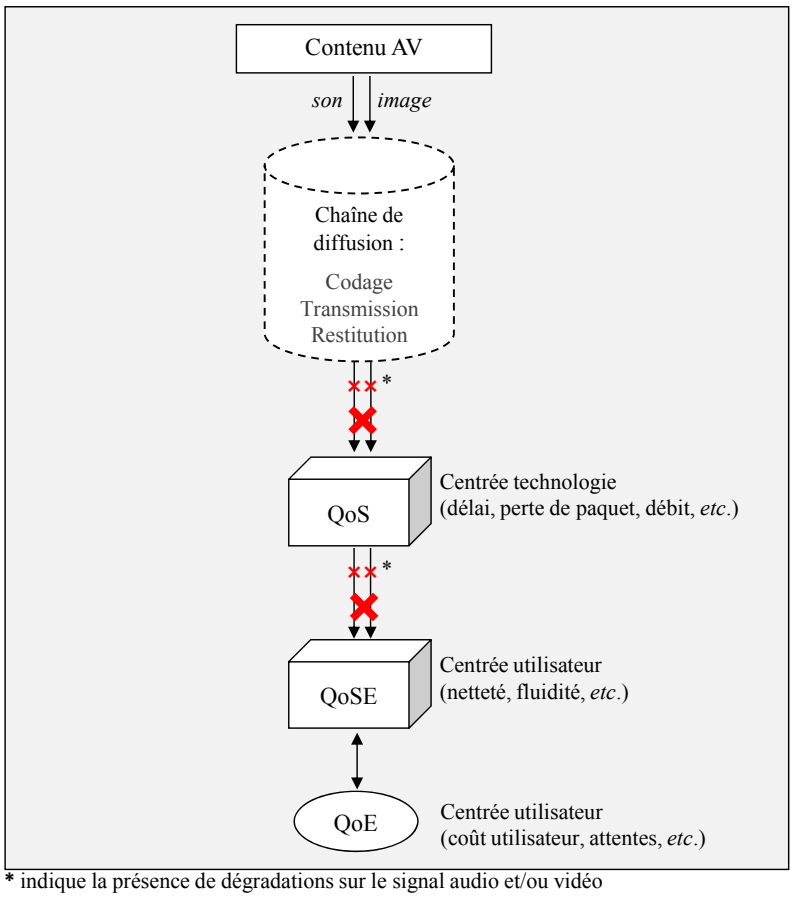
La qualité perçue du signal est désignée par le terme de QoSE (QoS E*xperienced*) et définie par la norme E.800 (2008) fournie par l'UIT (Union Internationale des Télécommunications) comme le niveau de qualité dont les utilisateurs estiment avoir bénéficié. Cependant, l'acceptabilité globale d'un service ne dépend pas seulement de la qualité perçue du signal (QoSE) mais également de caractéristiques liées au service (disponibilité, ergonomie, prix), au contexte d'évaluation ou encore à des aspects propres à l'utilisateur comme ses attentes vis-à-vis de ce service (expériences passées, but fixé, au contexte d'utilisation, *etc*.).

Le terme de QoE (Quality of Experience) ou *qualité d'expérience* est alors employé pour qualifier l'acceptabilité globale d'une application ou d'un service tel que subjectivement perçue par l'utilisateur final (UIT-T P.10/G.100, 2008). La QoE pourrait dépendre du coût pour l'utilisateur au sens de la fatigue ou de l'effort induit par une qualité dégradée des signaux restitués. Par exemple, de la fatigue ou une gêne visuelle peut être ressentie au cours ou après la visualisation de contenus AV présentant un format 3D vidéo (Chen, 2012 ; Lambooij, Fortuin, Heynderickx et IJsselsteijn, 2009 ; Ukai et Howarth, 2008 ; Yano, Ide, Mitsuhashi et Thwaites, 2002). La présence de fatigue ou de gêne pourrait réduire le confort ou la satisfaction de l'utilisateur voire conduire à un rejet du service en question. La mesure de la QoE, aux frontières de plusieurs disciplines comme la psychologie, les sciences cognitives ou les sciences de l'ingénieur, doit permettre de mieux comprendre les attentes et le ressenti de l'utilisateur vis-à-vis d'un système ou d'une technologie.

Ainsi, des dégradations n'auront pas la même influence sur la qualité selon qu'on la considère à travers un ensemble d'attributs techniques (performances du système, service),

perceptifs (rupture continuité, netteté, clarté : flou, saccades, distorsions) ou propres à l'individu (attentes de l'utilisateur, contexte d'utilisation, *coût utilisateur*, *etc.*).

Afin d'optimiser les services audiovisuels existants ou innovants et de garantir une qualité optimale à l'utilisateur, l'évaluation de la QoSE (qualité perçue) et plus largement de la QoE (*qualité d'expérience*) est un enjeu majeur pour les différents acteurs du domaine. La Figure 1.3 ci-dessous illustre le cheminement du signal audiovisuel de sa production à la *qualité d'expérience* finale pour l'utilisateur.



* indique la présence de dégradations sur le signal audio et/ou vidéo

Fig. 1.3. Cheminement du signal audiovisuel de sa production à la *qualité d'expérience* de l'utilisateur.

1.5. EVALUATION DE QUALITE D'EXPERIENCE : LES METHODES ACTUELLES

La *qualité d'expérience* est actuellement principalement étudiée à travers la qualité perçue (QoSE) des signaux audio et vidéo restitués en bout de chaine. Les méthodes actuelles d'évaluation de qualité utilisées par les instituts de télécommunications sont principalement recommandées par l'UIT. Un de ses objectifs est de mettre au point un ensemble de recommandations normatives, internationalement reconnu, permettant une comparaison inter-instituts des résultats issus des évaluations effectuées dans différents laboratoires. L'UIT établit donc un cadre commun d'évaluation définissant par exemple le contexte (conditions de visualisation et d'écoute : luminosité, niveau d'écoute, distance de visualisation, *etc.*) ou le

protocole à utiliser (échelles de jugement, nombre minimum de participants, durée et type des séquences de test, conditions à comparer, *etc.*).

Pour évaluer la qualité perçue des médias (QoSE) dans le domaine des télécommunications, deux types de méthodes sont considérés : les méthodes dites « objectives » ou instrumentales et les méthodes subjectives. Les premières vont s'appuyer sur des éléments de QoS (perte de paquets, délai) et/ou sur une analyse du signal pour prédire une note de qualité (prédiction QoSE). Les modèles utilisés pour les méthodes « objectives » sont souvent basés sur la mesure dite « subjective » ou perceptive. Ces mesures subjectives sont obtenues auprès d'utilisateurs à qui l'on demande d'évaluer la qualité de séquences ou la qualité de communications audio (téléphonie, téléconférence) ou audiovisuelles (visioconférence) réalisées avec un ou plusieurs autres testeurs. Les évaluations sont effectuées sur une ou plusieurs échelles de catégories.

En effet, bien qu'il soit largement admis que la qualité perçue des médias soit un phénomène multidimensionnel, la grande majorité des méthodologies d'évaluation font l'hypothèse que la qualité d'un signal audio et/ou vidéo peut être décrite par un scalaire sur une échelle unidimensionnelle de qualité. La notion de qualité est alors ramenée à une impression générale ou qualité globale, intégrant toutes les dimensions sous-jacentes.

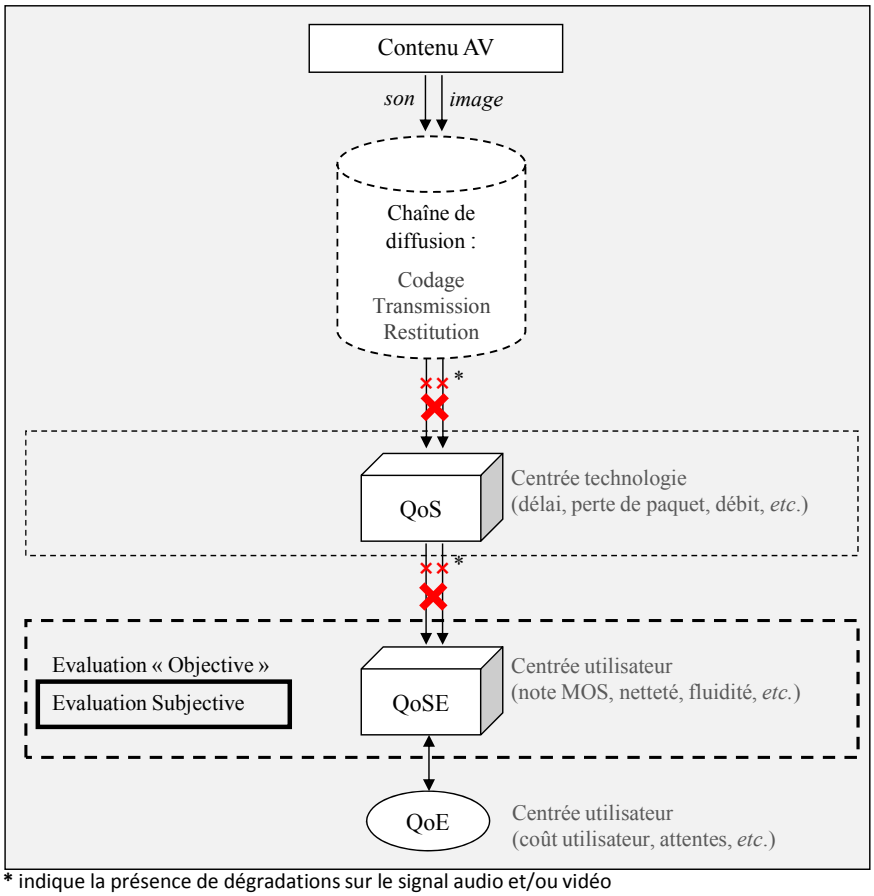
Les notes recueillies pour chaque individu sont ensuite moyennées, pour une séquence de test donnée, sur l'ensemble des participants. Le score moyen d'opinion ou le MOS (Mean Opinion Score) obtenu déterminera alors le niveau de qualité du signal évalué.

En règle générale, ces recommandations se concentrent sur l'évaluation d'une seule modalité, audio ou vidéo, à la fois. La norme UIT-T P.800 (UIT, 1996) est par exemple recommandée pour l'évaluation de la qualité vocale, les recommandations UIT-R BS.1284-1 (UIT, 2003), UIT-R BS.1534-1 (UIT, 2003) et UIT-R BS.1116-1 (UIT, 1997) permettent d'évaluer la qualité audio tandis que les normes UIT-R BT.500-13 (UIT, 2012) et UIT-R BT.1788 (UIT, 2007) sont dédiées à l'évaluation de la qualité vidéo. La BT.500-13 fournit par exemple des recommandations, entre autre, sur le contraste ou le rapport de luminance d'un écran. Certaines normes suggèrent également des méthodes pour évaluer une modalité donnée (audio ou vidéo) dans un contexte audiovisuel : la norme UIT-R BS.775-3 (UIT, 2012) et la norme UIT-R BS.1286 (UIT, 1997) permettent d'évaluer respectivement l'audio multicanal (radiodiffusion télévisuelle numérique) et les systèmes audio, de manière générale, en présence d'une image d'accompagnement et la norme UIT-T P.910 (UIT, 1999) propose des méthodes pour évaluer la qualité vidéo dans le cadre de systèmes multimédias tels que de la visioconférence. Seulement deux normes sont dédiées à l'évaluation subjective de la qualité audiovisuelle pour un contexte interactif (UIT-T P.920, 2000) ou non interactif (UIT-T P.911, 1998).

La norme UIT-T P.920 propose des recommandations pour l'évaluation de services de communication audiovisuelle (applications multimédias interactives comme la visioconférence). Les tâches de communication proposées (> 5 min) doivent inciter les participants à communiquer de la façon la plus naturelle possible et à rester concentrés sur le

média audiovisuel. La norme UIT-T P.920 décrit différents prétextes de communication pour engager le participant dans l'activité : jeu de questions/réponses, comparaison d'histoires ou d'images, *etc*. L'évaluation de la qualité audiovisuelle est réalisée à partir d'une approche multicritère. Il est notamment possible de demander aux participants de juger la qualité audiovisuelle globale mais également les qualités audio et vidéo jugées séparément. La norme UIT-T P.920 ne sera pas présentée plus en détails, son application étant dédiée à l'évaluation des systèmes multimédias interactifs.

La norme UIT-T P.911 propose des méthodes d'évaluation de la qualité audiovisuelle (AV) pour applications multimédias non interactives (contexte passif d'écoute et de visualisation : TV, multimédias, *etc.*). Le jugement de qualité s'effectue sur une seule et unique échelle à l'issue de la visualisation et de l'écoute de chaque séquence audiovisuelle de test. Quatre méthodes sont proposées dans le cadre de cette norme ; elles sont décrites dans les paragraphes suivants. La Figure 1.4 ci-après indique le niveau, dans le cheminement du signal, auquel interviennent les méthodes d'évaluations de la qualité perçue.



* indique la présence de dégradations sur le signal audio et/ou vidéo

Fig. 1.4. Cheminement du signal audiovisuel de sa production à la *qualité d'expérience* de l'utilisateur vis-à-vis du service audiovisuel utilisé.

1.5.1. MÉTHODE ACR : ABSOLUTE CATEGORY RATING

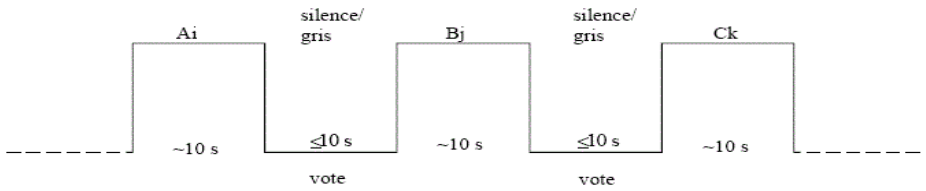
La méthode ACR ou méthode d'évaluation par catégories absolues consiste à attribuer une note de qualité après chaque séquence AV visualisée/entendue. La note de jugement attribuée doit refléter l'opinion du participant quant à la qualité audiovisuelle globale perçue, c'est-à-dire la qualité audio et vidéo combinée. Cette évaluation est réalisée sur une échelle catégorielle de cinq ou neuf points (intervalles) explicitée par cinq items (*Excellent-Bon-Satisfaisant-Médiocre-Mauvais*). Il est recommandé d'utiliser l'échelle en neuf points lorsqu'une plus grande puissance de discrimination est nécessaire, typiquement, lorsque l'on souhaite évaluer des codages à bas débit (UIT-T, 1999). Une illustration des échelles recommandées est apportée par la Figure 1.5.

9	Excellent
8	
7	Bon
6	
5	Satisfaisant
4	
3	Médiocre
2	
1	Mauvais

5	Excellent
4	Bon
3	Satisfaisant
2	Médiocre
1	Mauvais

Fig. 1.5. Echelle d'évaluation de qualité à 9 et 5 niveaux.

La norme recommande des séquences d'une durée comprise entre huit et dix secondes, l'intervalle de temps conseillé pour le vote est égal ou inférieur à dix secondes. Le chronogramme recommandé par la norme UIT-T P.911 est présenté par la Figure 1.6 ci-dessous.



Ai séquence A dans la condition d'essai i
Bj séquence B dans la condition d'essai j
Ck séquence C dans la condition d'essai k

Fig. 10.6. Chronogramme de la méthode ACR (issu de la P.911, UIT-T, 1998).

La méthode ACR est une méthode peu coûteuse du point de vue de son application, du traitement et de l'analyse des résultats. Elle présente également l'avantage de pouvoir qualifier des systèmes à l'essai et d'obtenir leur hiérarchisation selon le niveau de qualité leur étant associé.

1.5.2. MÉTHODE DCR : DEGRADATION CATEGORY RATING

La méthode DCR ou méthode par évaluation de catégories de dégradations propose une présentation des séquences AV de test par paires. Les séquences constituant la paire sont identiques à la différence que la première est toujours présentée sans dégradations (référence) tandis que la seconde est traitée par le système à évaluer (donc susceptible de comporter des dégradations). La séquence traitée est toujours présentée après la référence.

Seule la séquence traitée est évaluée par les participants en comparaison avec la condition de référence. L'échelle d'évaluation correspond ici à une échelle de perceptibilité de la dégradation comme présenté par la Figure 1.7.

Les durées des séquences de test et du temps de vote sont identiques à celles recommandées dans la méthode ACR. L'avantage principal de cette méthode est de permettre une qualification rapide du niveau de gêne associé à certaines dégradations engendrées par les systèmes considérés.

5	Imperceptible
4	Perceptible mais non gênant
3	Légèrement gênant
2	Gênante
1	Très gênante

Fig. 1.7. Echelle de dégradation à cinq niveaux.

1.5.3. METHODE PC : PAIR COMPARISON

La méthode PC ou par comparaison de paires consiste à présenter deux séquences identiques, à la différence que chaque séquence est traitée par un système à l'essai différent. La séquence de référence (sans dégradations) peut également être incluse en tant que système à l'essai additionnel. Toutes les combinaisons de paires de séquences A, B, C, *etc.* devront être évaluées (AB, BA, CA, *etc*.) et présentées dans les deux ordres possibles (AB, BA, *etc*.). Le jugement de qualité AV globale est ici exprimé à travers un jugement de préférence pour l'une ou l'autre séquence de la paire. Ce jugement est réalisé après la présentation de chaque paire. Cette méthode est notamment préconisée pour la comparaison de systèmes quasi-équivalents et/ou de haute qualité. La durée recommandée pour les séquences de test est d'environ dix secondes, celle du temps de vote doit être inférieure ou égale à dix secondes.

1.5.4. METHODE SSCQE : SINGLE-STIMULUS CONTINUOUS QUALITY EVALUATION

Une dernière méthode, la méthode SSCQE (méthode d'évaluation continue de la qualité avec stimulus unique) est une méthode de jugement en continu permettant de recueillir l'évaluation des participants pendant la visualisation des séquences de test pour lesquelles le niveau de qualité fluctue. Les testeurs reportent leur jugement au moyen d'un curseur pouvant être déplacé le long d'une échelle continue. Celle-ci permet d'attribuer une note entre 0 et 100 où 100 représente une qualité parfaite. L'échelle est divisée en cinq segments égaux qui

correspondent à l'échelle de qualité à cinq points, les items caractérisant les différents niveaux sont identiques à ceux de la méthode ACR.

Aucune référence n'est donnée pour servir de base à l'évaluation subjective. La durée de séquences de test proposées est beaucoup plus importante que les précédentes méthodes. Celle-ci peut en effet être comprise entre trois et trente minutes.

La compression du signal de télévision numérique va entraîner des dégradations pour la qualité de l'image et du son, dégradations qui dépendent de la scène et varient en fonction du temps. Cette méthode permet l'observation de l'influence des fluctuations de qualité (plus réalistes qu'une application constante des dégradations sur une séquence donnée), au moment où elles se produisent, sur la note de jugement du participant. En d'autres termes, la méthode SSCQE présente l'avantage de pouvoir étudier l'impact des variations de qualité sur la perception du participant en temps réel. De plus, l'absence de référence et l'allongement de la durée de visualisation propose un contexte d'évaluation plus proches de conditions réelles de visualisation.

Le choix d'une méthode par rapport à une autre sera guidé selon que l'objectif fixé corresponde à une discrimination fine entre plusieurs systèmes, à une qualification de systèmes ou encore à une détection de dégradations.

Ces méthodes présentent plusieurs avantages tels que leur facilité d'application, de traitement ou d'analyse des résultats. Les notes MOS collectées grâce aux méthodes proposées par la norme UIT-T P.911 permettent de hiérarchiser rapidement plusieurs systèmes en matière de qualité perçue, de gêne liée à la présence de dégradations ou encore de préférence.

Cependant, celles-ci présentent un certain nombre de biais et de faiblesses auxquels il faut prêter attention lors de l'interprétation des données récoltées.

1.6. FAIBLESSES DES METHODES ACTUELLES

Les méthodes proposées par la norme UIT-T P.911 comportent des biais inhérents à la mesure subjective.

1.6.1. BIAIS DE REPRESENTATIVITE

Une faiblesse majeure des méthodes actuelles est de limiter l'étude de la QoE à la seule évaluation de la qualité perçue du signal audio et/ou vidéo restitué. Les notes recueille ne reflètent pas un certain nombre d'effets potentiellement induit par la perte de qualité comme, par exemple, des effets liés à la présence d'un effort ou de fatigue. En conséquence, la représentativité de la QoE à travers les notes de qualité est restreinte.

Par ailleurs, le processus de jugement de qualité, demandé et effectué de manière explicite, est intrusif, et peu représentatif de l'utilisation de services dans la vie quotidienne des

utilisateurs (Gros *et al*., 2006). En effet, l'évaluation consciente de qualité constitue un biais en soi puisqu'elle focalise l'attention de l'observateur sur le niveau de qualité. Cette approche est peu représentative d'un contexte réel d'utilisation pour lequel l'utilisateur n'a souvent pas conscience de ces aspects de qualité hormis en présence de dégradations importantes. La qualité des médias n'est donc généralement pas un objet conscient.

La courte durée (entre 8 et 10 s) des séquences de test, recommandée pour éviter l'influence d'effets mnésiques (primauté et récence), est également peu représentative des durées réelles de visualisation lors de l'utilisation de services AV. Or, un niveau de qualité perçue comme suffisant pour une séquence de courte durée pourrait diminuer lors d'une exposition plus longue. En effet, si dix secondes suffisent pour rendre les dégradations perceptibles et à même d'être jugées (en 1999, Vahedian, Frater et Arnold ont montré que 5 s suffisent pour permettre aux participants d'émettre un jugement de qualité), elle ne permet pas d'étudier l'impact de dégradations sur du long terme. Par exemple, des dégradations récurrentes durant la visualisation et l'écoute d'un contenu comme une désynchronisation image/son ou la présence de saccades vidéo pourraient conduire à un désagrément important qu'une séquence de dix secondes ne permettrait d'observer. La présence de dégradations sur du plus long terme pourrait également induire des phénomènes de fatigue ou d'effort pouvant être à l'origine d'un rejet du système. Cette question a notamment été soulevée dans le cadre de diffusion au format 3D vidéo pour lequel des états de fatigue visuelle peuvent être ressentis. La norme UIT-T P.911 propose cependant d'utiliser des séquences plus longues dans le cadre de la méthode SSCQE. Toutefois, le contexte proposé est peu propice au maintien de l'intégrité de l'activité de jugement ou de visualisation. En effet, la « tâche » d'évaluation en continu implique un déplacement attentionnel entre ces deux activités, c'est-à-dire que le participant devra momentanément se détacher du contenu audiovisuel pour déplacer le curseur. Cette méthode s'éloigne ainsi de conditions réelles d'utilisation et son application même peut constituer un biais dans le jugement des participants (détachement puis réengagement dans l'activité de visualisation).

L'évaluation subjective soulève également deux principaux biais de représentativité : un biais de rationalisation (comparaison inter-séquences et/ou jugement réalisée au moyen d'une référence implicite spécifique ou extérieure au test et propre à un participant donné) et un biais de conformité aux attentes de l'expérimentateur (réflexe de la bonne réponse). Ce dernier relève d'un automatisme, conscient ou non, de désirabilité sociale où l'individu présente une tendance à donner une description positive de lui-même (Paulhus, 2002), c'est-à-dire que la réponse est souvent dirigée vers celle supposée comme étant attendue.

1.6.2. BIAIS DE LA MESURE

Les outils mêmes utilisés pour recueillir les jugements des participants constituent une limite à la validité des méthodes UIT (Gros, 2001). En effet, les échelles de catégories utilisées peuvent fausser le jugement en proposant un champ trop restreint (ne recouvrant pas

toutes les situations) ou trop large d'items (au-delà de 5 ou 6 catégories, le participant ne serait plus capable de catégoriser les stimuli sans confusion).

De plus, comme le soulèvent Mullin, Smallwood, Watson et Wilson G. M. (2001), une autre difficulté peut provenir de l'uniformisation des items recommandés. En effet, l'objectif de l'UIT est de proposer des échelles pour une utilisation standardisée à un niveau international. Cependant, la traduction des items est généralement littérale et n'est validée par aucune étude préalable. Les différences d'interprétation des items et l'écart entre ces items selon la composante culturelle ou linguistique n'est pas considérée. Mullin *et al.* soulignent pourtant que l'interprétation des items est soumise à une grande variabilité culturelle. Par exemple, le terme « OK » a une signification différente pour des américains, qui le considèrent comme équivalent au terme anglophone « Fair » et pour des italiens qui l'assimile plutôt au terme « Good » (Jones et McManus, 1986, cité par Mullin *et al.*). Diverses études ont également montré que la distance entre les intervalles de catégories n'est pas forcément régulière et identique d'une culture à l'autre. Mullin *et al.* décrivent une étude impliquant vingt-quatre participants anglophones devant positionner un certain nombre d'items, dont ceux proposés par l'UIT (*Excellent-Good-Fair-Poor-Bad*), sur une ligne de 200 mm. Les annotations réalisées permettaient d'obtenir les écarts de perception entre les items évalués. Les résultats obtenus ont clairement montré que les items ne sont pas représentés par des intervalles équivalents comme l'illustre la Figure 1.8 ci-après. Ce constat tend à être confirmé par Jones et McManus (1986, cité par Mullin *et al.*). Ces auteurs ont montré que les items anglophones « Bad » et « Poor » présentent une distance perceptive très faible tandis que cette distance est beaucoup plus importante entre les items « Good » et « Fair ». L'absence de régularités entre intervalles a également été observée lors de l'évaluation de ces items en langue suédoise (Virtanen, Gleiss et Goldstein, 1995, cité par Mullin *et al.*). Les équivalents des termes « Bad » et « Poor » étaient perçus comme étant très similaires tandis qu'un écart important a été obtenu entre « Poor » et «Fair ».

Ainsi, les échelles catégorielles et les items explicitant ces catégories sont soumis à la multiplicité individuelle et culturelle de leur interprétation pouvant remettre en cause la validité des notes de qualité recueillies.

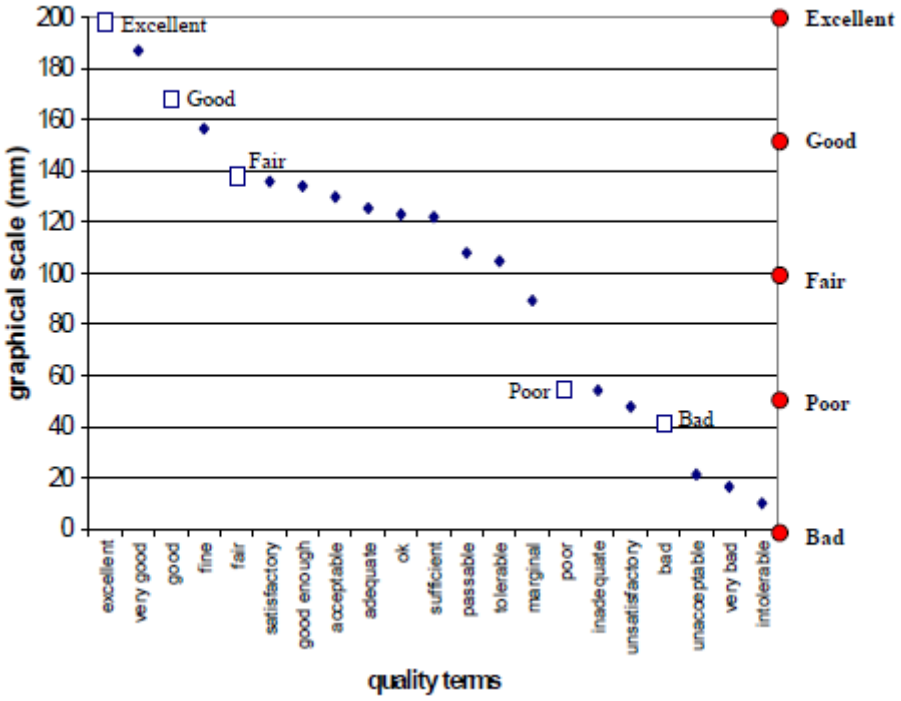


Fig. 1.8. Positions moyennes des items UIT. L'axe de droite présente les positions théoriques des items, c'est-à-dire telles que proposées par les échelles UIT.

Par ailleurs, les notes MOS collectées sont dépendantes du corpus de séquences de test évaluées. En effet, le testeur va moins juger le niveau de qualité dans l'absolu que procéder par comparaison en s'appuyant sur les autres séquences. Une même condition de qualité pourra ainsi être jugée de bonne qualité dans un corpus constitué de conditions de mauvaise qualité mais de qualité moyenne dans un corpus constitué de conditions de bonne qualité.

Ces conditions de qualité sont appliquées à des séquences de test (échantillons de parole, de musique, extraits vidéo ou audiovisuels) dont les caractéristiques participent fortement à l'élaboration du jugement de qualité. Le choix de ces séquences est alors crucial en raison des biais qu'il peut induire sur la mesure subjective recueillie.

1.6.3. BIAIS DU CONTENU AUDIO ET VIDEO

Un aspect négligé par la norme UIT-T P.911 concerne l'influence du contenu sémantique des séquences de test sur la perception de qualité. Dans ce document, la notion de sémantique se réfère à l'ensemble des éléments indispensables pour construire le sens de la scène. La norme UIT-T P.911 propose une caractérisation du contenu audio et vidéo des séquences de test mais considérée séparément, c'est-à-dire que le lien entre les deux modalités n'est pas pris en compte. En d'autres termes, le contenu AV n'est pas décrit comme un évènement global qui tiendrait compte du rapport entre son et image. Par exemple, la caractérisation ne précise pas la modalité véhiculant l'information principale (une dégradation survenant sur cette

modalité pourrait plus fortement altérée la qualité perçue) ou encore la relation entre l'audio et la vidéo (son d'une scène en lien ou non avec la vidéo par exemple).

De plus, les caractéristiques proposées apportent peu d'informations en matière de description technique (luminosité, température de couleur, dynamique caméra, *etc*.) et sémantique (dynamique perçue, compréhension, *etc*.) qui pourrait interagir avec la qualité perçue.

De la même manière, il est envisageable que la qualité hédonique d'une séquence (plaisir, intérêt, *etc*.) puisse diminuer ou augmenter la tolérance de l'utilisateur vis-à-vis de certaines dégradations. Par exemple, Palhais, Cruz et Nunes (2012) ont montré que le niveau d'intérêt a une influence positive sur l'évaluation subjective de la qualité vidéo : lorsque le niveau d'intérêt augmente, la note de qualité augmente également.

Ainsi, selon la séquence visualisée, des dégradations identiques pourraient ne pas avoir la même influence sur le niveau de qualité perçue, en raison de caractéristiques techniques plus ou moins bien gérées par les systèmes à évaluer (un codeur vidéo donné gérera plus ou moins bien le mouvement ou les changements de luminosité par exemple, de par les choix faits au moment de son implémentation), de l'intérêt ou des émotions suscités par le contenu même de la séquence et/ou de la sémantique du contenu.

Le Tableau 1.1 ci-après récapitule les spécificités, avantages et inconvénients principaux de chaque méthode d'évaluation (ACR, DCR, PC ou SSCQE) recommandée par la norme UIT-T P.911.

Tableau 1.1. Synthèse des méthodes de recueil du jugement de qualité perçue et de leurs spécificités.

<table>
<tr><td></td><td>Référence explicite</td><td colspan="3">Sans référence explicite</td></tr>
<tr><td></td><td>DCR</td><td>ACR</td><td>PC</td><td>SSCQE</td></tr>
<tr><td>Stimuli</td><td>Paires :
référence/dégradation</td><td>Stimulus unique</td><td>Paires :
2 systèmes à l'essai</td><td>Stimulus unique</td></tr>
<tr><td>Durée</td><td>[8 - 10 s]</td><td>[8 - 10 s]</td><td>[8 - 10 s]</td><td>[3 - 30 min]</td></tr>
<tr><td>Evaluation</td><td>A posteriori</td><td>A posteriori</td><td>A posteriori</td><td>Continue</td></tr>
<tr><td>Echelles</td><td>Catégorielle
(5 points) :
[Imperceptible - très gênante]</td><td>Catégorielle
(5/9 points) :
[Excellent-mauvais]</td><td>Préférence pour l'une des deux séquences de la paire</td><td>Continue 0-100
(5 points) :
[Excellent-mauvais]</td></tr>
<tr><td>Avantages</td><td>Pour système de haute qualité

Détection des dégradations</td><td>Qualification de système

Facile, application rapide</td><td>Haut pouvoir de discrimination

Utile pour comparer 2 systèmes d'essai de qualités quasi-égales</td><td>Etude des fluctuations de qualité en temps réel</td></tr>
<tr><td rowspan="2">Faiblesses</td><td colspan="3">Séquences de 10 s
Evaluation a posteriori</td><td>peu adaptée à une situation de visualisation</td></tr>
<tr><td colspan="4">Evaluation de la qualité explicite
Biais de représentativité
Biais de la mesure</td></tr>
</table>

1.7. VERS UNE SOLUTION ALTERNATIVE

En résumé, les méthodes subjectives recommandées par l'UIT comportent un certain nombre de biais limitant la mesure de qualité et son interprétation. Le protocole même de recueil des notes de qualité (échelles, items, durées des séquences, *etc.*) diminue la représentativité et la fiabilité de la mesure. Au-delà des biais propres à la mesure subjective, les méthodes actuelles restreignent l'évaluation de la *qualité d'expérience* (QoE) à l'évaluation de la qualité perçue des signaux audio et vidéo restitués (QoSE). Or une qualité dégradée pourrait influencer plus largement la *qualité d'expérience* de l'utilisateur par exemple en induisant de la fatigue ou un effort mental (*coût utilisateur*). Ces aspects peuvent notamment être étudiés à travers des mesures physiologiques et oculaires qui présentent l'avantage de ne pas être soumises aux biais inhérents à la mesure subjective.

Plusieurs études ont montré que des dégradations de qualité peuvent influencer l'activité physiologique des utilisateurs (Wilson G. M. et Sasse, 2000a, 2000b, ces études seront détaillées dans le chap. V). La qualité est envisagée par ces auteurs sous l'angle du coût pour l'utilisateur du point de vue de l'effort fourni pour accéder au sens du message audio et/ou vidéo. Le coût induit peut être étudié au moyen de mesures de l'activité physiologique telles que le rythme cardiaque ou la sudation cutanée. L'étude du comportement oculaire (durée et fréquence des clignements des yeux, *etc.*) peut également permettre d'observer des phénomènes comme l'effort mental ou la fatigue.

Ainsi, une méthode alternative pour étudier l'influence de la qualité des signaux audio et vidéo restitués sur la qualité de l'expérience du spectateur pourrait consister à recueillir conjointement des mesures subjectives ainsi que des mesures de l'activité physiologique et oculaire des spectateurs en situation de visualisation et d'écoute de séquences audiovisuelles 2D ou 3D. C'est dans cette direction que ce travail de recherche s'inscrit comme cela sera exposé dans les chapitres suivants.

CHAPITRE II – PERCEPTION BIMODALE AUDIOVISUELLE

La perception de la qualité audiovisuelle relève des capacités du système audiovisuel humain (SAVH) à traiter et à interpréter l'information audiovisuelle. En effet, l'influence d'une dégradation sur le niveau de qualité perçue sera plus ou moins importante selon la sensibilité du SAVH à la percevoir. La première partie de ce chapitre portera sur le processus d'intégration des informations auditives et visuelles pour aboutir à leur fusion et former le percept audiovisuel global. Cette étape permettra de mieux comprendre les principes sous-jacents à la perception bimodale. Les disparités temporelles et spatiales capables d'interférer voire d'empêcher ce processus d'intégration audiovisuelle (et à terme d'entraîner une diminution de la qualité perçue) seront abordées.

Une seconde partie sera dédiée à l'étude des influences mutuelles entre qualité audio et vidéo ainsi que leur contribution respective à la qualité audiovisuelle globale perçue par l'utilisateur. Les différents facteurs (contenu, attention, *etc*.) pouvant influencer la perception de la qualité seront également traitées au cours de cette seconde partie.

2.1. PERCEPTION AUDIOVISUELLE

La perception est généralement décrite comme une fenêtre multisensorielle (visuelle, auditive, tactile, olfactive, gustative) ouverte sur le monde extérieur. Dans le cas de la perception bimodale audiovisuelle, les informations capturées par les systèmes sensoriels visuels et auditifs humains vont être acheminées et traitées par les différents systèmes perceptifs dédiés. La principale fonction des organes sensoriels auditifs et visuels est de capter les stimulations physiques de l'environnement. Le système visuel répond principalement à des longueurs d'ondes spécifiques du spectre lumineux tandis que le système auditif capte une plage particulière de fréquences sonores. Une seconde fonction réside dans leur capacité à transformer cette stimulation en influx nerveux grâce à la transduction du signal physique en potentiels d'actions. L'énergie lumineuse est absorbée par les cellules de la rétine et transformée en influx nerveux par un processus de photo-transduction s'opérant dans les cônes et les bâtonnets. Les ondes sonores sont traduites en mouvements mécaniques transmis de la membrane du tympan à la cochlée où l'information acoustique est transformée en influx nerveux. L'information nerveuse obtenue est ensuite transmise aux différentes aires corticales *via* les nerfs optiques ou auditifs.

Pour être traitée, cette information nerveuse est parcellisée en différentes propriétés (primitives) traitées distinctement par différentes parties des systèmes visuels ou auditifs. Les primitives visuelles correspondent par exemple à la détection des contours, au contraste, à la couleur, au mouvement et à son orientation, à la forme ou encore aux aspects stéréoscopiques de l'information visuelle. Le signal auditif entrant est également morcelé en différents attributs tels que la hauteur, la sonie, le timbre, la position, *etc*.

L'information nerveuse visuelle ou auditive est donc traitée de manière parcellaire (forme, couleur, mouvement, localisation, intensité, *etc*.) et parallèle dans les différentes aires corticales. L'intégration synchrone des informations issues des différentes aires permettra d'aboutir à la perception visuelle ou auditive finale. La reconnaissance et l'interprétation des

informations feront appel à des composantes attentionnelles/émotionnelles et seront mises en relation avec les connaissances antérieures stockées en mémoire.

La description fonctionnelle succincte du précédent paragraphe (une présentation plus amplement détaillée des bases fonctionnelles et anatomophysiologiques des systèmes auditifs et visuels humains est apportée dans l'annexe 2-A) a essentiellement présenté les deux systèmes de traitement des informations auditives et visuelles comme deux modules étanches. Pourtant, les différentes informations sensorielles ont souvent besoin d'être intégrées simultanément pour rendre possible une interprétation compréhensible du monde extérieur. Par exemple, au cours d'une discussion, la source auditive est généralement identifiée comme provenant du locuteur. Ce lien associatif est permis grâce à une convergence des informations auditives et visuelles. Sans cette convergence, les informations audiovisuelles locuteur/paroles seraient dissociées ce qui rendrait difficile, voire impossible, l'obtention d'une interprétation correcte de la situation.

2.1.1. INTEGRATION MULTIMODALE

La plupart des évènements issus de notre environnement fournissent des informations *via* les différentes modalités sensorielles. Ces informations sont généralement intégrées dans une représentation multisensorielle unique (King, 2005; Spence, 2007; Stein et Meredith, 1993). Kohlrausch et van de Par (1999) ont défini l'intégration multisensorielle comme la synthèse d'informations issue d'au moins deux modalités sensorielles et donnant lieu à une information émergente ou à une signification (percept unique). Le sens extrait de cette synthèse n'aurait pu être obtenu par un traitement séparé des informations issues de chacune de ces modalités. Les auteurs illustrent ce phénomène par l'expérience familière que représente l'action de sonner à une porte : la vision de l'objet, la stimulation tactile du déclenchement de la sonnette et le son conséquent sont facilement intégrés en un unique évènement indiquant ainsi le bon fonctionnement de la sonnette et la réussite de l'action initiée. Cette intégration multisensorielle serait déjà fonctionnelle chez le nourrisson (Bahrick, 1987), ce qui révèle l'importance de ce groupement multimodal pour capter et apprendre des régularités de l'environnement à partir des différentes modalités sensorielles. Cette unité permet de s'adapter à l'environnement par une interprétation adéquate des évènements perçus.

Cette intégration peut intervenir soit à un niveau précoce (Foxe et Schroeder, 2005), c'est-à-dire qu'elle survient au sein même des aires sensorielles, soit plus tardivement au sein d'aires intégratives de plus hauts niveaux. Ces zones intégratives hébergeraient des neurones dits multisensoriels avec pour particularité de pouvoir répondre à des stimuli issus de modalités différentes : auditive, visuelle, somatique, *etc.* Divers sites d'intégration multisensorielle existent (zones corticales préfrontales, pariétales ou temporales et sous-corticales comme les noyaux gris centraux ou le thalamus) mais le principal serait localisé dans le colliculus supérieur décrit dans le paragraphe suivant.

SUBSTRATS NEURONAUX DE L'INTEGRATION MULTIMODALE

Les colliculi ou tubercules quadrijumeaux sont des structures sous-corticales appartenant au tronc cérébral. Deux types de colliculi peuvent être identifiés : les colliculi inférieurs et les colliculi supérieurs. Les premiers sont impliqués dans le traitement auditif notamment de localisation (Wagner, 1993) alors que les seconds sont principalement responsables des réflexes d'orientation visuelle (Koch, 2004).

Cependant, les colliculi inférieurs (CI) sont également capables de répondre à différentes entrées sensorielles. En effet, en plus des entrées auditives, les CI reçoivent aussi des entrées somatosensorielles (Paloff et Usunoff, 1992) et visuelles (Mascetti et Strozzi, 1988). Par ailleurs, Groh, Trause, Underhill, Clark et Inati (2001) ont montré que les mouvements oculaires modulent les réponses des neurones auditifs du colliculus inférieur chez le primate. En d'autres termes, la position des yeux aurait une influence sur le traitement spatial de l'information auditive. Ce constat pourrait en partie permettre d'expliquer certaines influences des interactions entre les modalités auditives et visuelles dont l'impact comportemental sera abordé dans la section 2.1.2 ci-après.

Malgré la fonction intégrative du CI, l'intégration multimodale est généralement rapprochée du colliculus supérieur (CS). Des études neurophysiologiques (pour une revue voir Stein et Meredith, 1993) ont en effet révélé l'existence de neurones multimodaux (bi ou tri-modaux) dans cette zone. Le CS se caractérise par une structure laminaire composée de sept couches de cellules différentes. Les couches les plus superficielles (couches dorsales) seraient exclusivement impliquées dans les traitements d'informations visuelles projetées de la rétine et du cortex visuel primaire. Les couches les plus profondes (couches ventrales), généralement associées aux afférences motrices, recevraient également des projections visuelles, auditives ou somesthésiques. Il semblerait donc que les informations auditives, visuelles et somatiques convergent vers cette zone. La particularité des neurones multimodaux est donc de recevoir des afférences d'informations extraites de plusieurs modalités d'un même objet ou d'un même évènement. Comme le montre la Figure 2.1 ci-dessous, la réponse des neurones multimodaux présente un accroissement multiplicatif -jusqu'à douze fois- de la décharge lors de la présentation de stimuli multimodaux comparativement à des entrées unimodales (c.-à-d. que la réponse multimodale est supérieure à la somme des réponses obtenues lors de stimuli unimodaux).

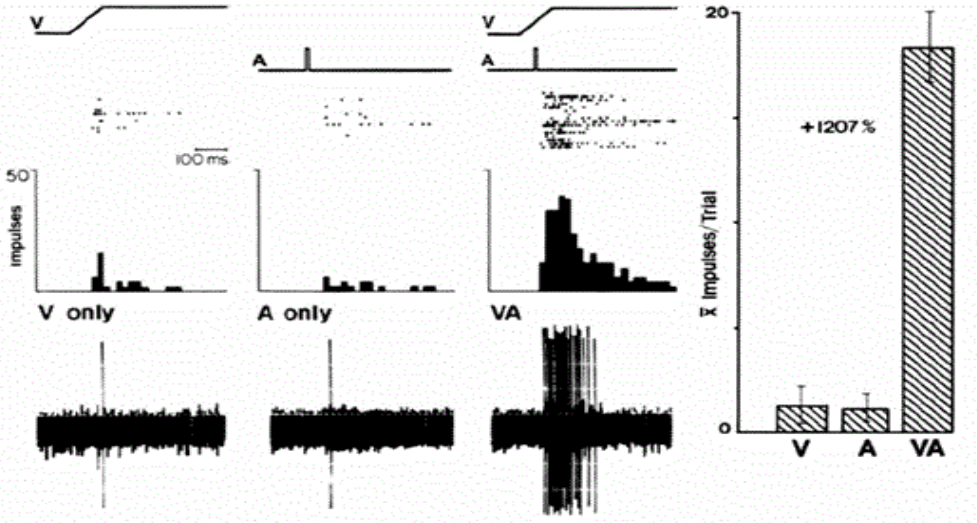


Fig. 2.1. Accroissement de la réponse audiovisuelle (VA) comparativement aux réponses unimodales auditives (A) et visuelles (V) (extrait de Stein et Meredith, 1993, p. 124).

Par ailleurs, plus les informations extraites sont appariées temporellement et spatialement, plus la réponse de ces neurones sera importante (Stein et Meredith, 1993). A l'inverse, en cas de disparités temporelles ou spatiales entre les différents indices sensoriels, la réponse sera moins importante. Ce constat indique que l'intégration multimodale va dépendre du respect de différents principes d'intégration au niveau du neurone unitaire (voir Stein et Meredith, 1993). Ces derniers correspondent aux principes de :

- ***préservation des champs récepteurs*** *:* les propriétés des champs récepteurs unimodaux sont préservées lors de l'intégration multisensorielle,
- ***efficacité inversée*** *:* plus un stimulus unimodal est efficace (saillance) moins la réponse obtenue lorsque que ce stimulus est combiné avec une autre modalité sera forte, et inversement. En d'autres termes, aucun gain n'est obtenu par l'ajout d'une modalité lorsqu'un stimulus unimodal est suffisamment efficace seul, à l'inverse, si ce stimulus unimodal est peu efficace (non saillant), sa réponse peut augmenter lorsque ce dernier est combiné avec une autre modalité,
- ***coïncidence spatiale*** : il y aura augmentation de la réponse multimodale si les sources unimodales sont issues de la même source spatiale, dans le cas contraire, la réponse diminue,
- ***coïncidence temporelle*** : il y aura augmentation de la réponse multimodale si les sources unimodales sont temporellement proches. Cependant, le délai de transduction des récepteurs sensoriels aux cellules du CS est dépendant du type de modalité. En effet, le délai de transduction est de l'ordre de 40 à 120 ms pour les stimuli visuels contre seulement 6 à 25 ms pour les stimuli audio (Stein et Meredith, 1993 ; Stein, Wallace et Meredith, 1995). En d'autres termes, le stimulus auditif est transmis plus rapidement que le signal visuel aux cellules du CS. Selon Arrighi, Alais, Burr (2006) et King (2005), la transduction du signal auditif serait plus rapide de 40 à 50 ms par rapport à l'information visuelle. Cet écart est

> généralement compensé par une activité neuronale unimodale suffisamment longue pour permettre un chevauchement des réponses issues des différentes modalités notamment grâce à une fenêtre d'intégration d'environ 1500 ms. Ces différences de délai de transduction viennent compenser les différences de célérité de propagation des ondes sonores et visuelles (~340 m/s *vs.* ~300 000 km/s respectivement). Cette disparité suppose notamment qu'il est physiquement impossible que le signal acoustique émis par une source visible soit perçu par l'observateur avant l'image de cette même source.

Le respect des règles de coïncidence temporelle et spatiale va donc permettre aux neurones multimodaux de répondre de façon optimale aux stimulations audiovisuelles. Ces principes de coïncidence s'étendent aux mesures comportementales pour lesquelles des stimuli unimodaux sont plus facilement détectés et identifiés comme provenant d'une même source lorsque leur origine spatiale et temporelle est commune. En situation de conflits (de localisation ou de synchronie) entre les modalités auditives et visuelles, le SAVH va tenter de maintenir, dans une certaine mesure, une représentation multimodale unique. Le maintien de cette unité malgré la présence de disparités suppose que les caractéristiques perçues d'une modalité sont modifiées par la présence de l'autre modalité. Par exemple, lors de disparités spatiales entre les sources auditives et visuelles, une modalité peut attirer dans sa direction la perception de l'autre modalité afin d'obtenir un évènement audiovisuel unique, on parle alors d'interactions audiovisuelles.

2.1.2. INTERACTIONS AUDIOVISUELLES

Welch et Warren (1989) définissent le phénomène d'interaction comme l'altération, dans le cas d'un percept bimodal, de la perception d'une modalité par la présence d'un stimulus sur l'autre modalité. Dans le cas de contenus audiovisuels, l'interaction se traduit par l'altération de l'audition par la vision ou inversement. D'un point de vue général, l'interaction bi ou multi -modale correspond aux influences mutuelles qu'exercent les modalités les unes sur les autres. Les paragraphes suivants décriront plus en détails les phénomènes d'interactions entre les modalités auditives et visuelles.

2.1.2.1. FUSION AUDIOVISUELLE

Deux notions très proches sont utilisées pour décrire les phénomènes d'interactions audiovisuelles : le pairing et la fusion.

PAIRING

Le *pairing* (Epstein, 1975) se définit comme le mécanisme sous-tendant les influences de bas-niveau qu'exercent, l'une sur l'autre, les modalités auditives et visuelles. Ces interactions de bas-niveau se produisent de manière automatique et non consciente, les facteurs conceptuels et attentionnels y joueraient un rôle peu influent (Radeau, 1994 ; Radeau et

Bertelson, 1977, 1978 ; Ragot, Cavé et Fano, 1988). Toutefois, le pairing est soumis aux principes de synchronie et de proximité spatiale et sa manifestation peut être envisagée comme la traduction comportementale des principes de coïncidence spatiale et temporelle à l'origine des réponses des neurones multimodaux du colliculus supérieur.

Par exemple, Bertelson et Radeau (1981) ont montré que la localisation d'un son ou d'un flash lumineux est biaisée lors d'une présentation simultanée conflictuelle (séparation angulaire de 7°, 15° ou 25°), par le stimulus de l'autre modalité. Plus précisément, la localisation d'un son sera attirée dans la direction du flash et *vice-versa*. Le biais observé témoigne de la tendance du système perceptif à toujours vouloir recréer un évènement audiovisuel cohérent même lorsque les indices perceptifs ne sont pas totalement appariés. Cependant, ce biais diminue à mesure que l'écart angulaire augmente (Bermant et Welch, 1976 ; Bertelson et Radeau, 1981).

L'étude de Bertelson et Radeau a également révélé une influence plus importante de la vision sur l'audition que la condition inverse. Ce constat pourrait être expliqué par un effet de consignes. En effet, selon Ragot *et al.* (1988), les consignes données aux participants (à savoir de porter leur attention sur l'une ou l'autre modalité) auraient une influence sur la prédominance d'une modalité sur l'autre et plus généralement sur les interactions de ces modalités. Toutefois, la prédominance de la vision sur l'audition pourrait également être mise en regard des spécificités propres à chaque système perceptif, visuel ou auditif, en matière d'acuité spatiale et temporelle. Le système visuel présente une résolution spatiale nettement meilleure que le système auditif, respectivement une minute d'arc (soit 1/60 de degré, Howard, 1982) contre un degré (Mills, 1958). En revanche l'acuité temporelle du système auditif (2 ms, Hirsh et Sherrick, 1961) est supérieure à celle du système visuel (20 ms, Haber et Hershenson, 1980). Partant de ce constat, Welch et Warren (1980) ont postulé que les processus multisensoriels répondent à l'hypothèse de la modalité pertinente. Celle-ci consiste à croire que l'une des modalités du percept multimodal serait prioritaire en raison de sa capacité à décrire un stimulus donné avec la meilleure précision (Welch, DuttonHurt et Warren, 1986). Cette hypothèse expliquerait la prédominance de la modalité visuelle pour les tâches spatiales et pour lesquelles le stimulus auditif serait fortement dépendant du stimulus visuel. Parallèlement, l'audition est considérée comme dominante pour les tâches de discriminations temporelles.

FUSION

La notion de *fusion perceptive* (Radeau et Bertelson, 1977) est utilisée pour désigner les phénomènes conscients d'intégration multimodale. La fusion serait influencée par des facteurs cognitifs tels que des facteurs attentionnels, conceptuels, *etc.* Par exemple, le phénomène de fusion audiovisuelle serait initié ou renforcé en présence d'un certain degré de cohérence entre les modalités audio et vidéo tel qu'un son de parole et l'image du locuteur (Jackson, 1953 ; Thurlow et Jack, 1973) ou autres évènements audiovisuels pour lesquels la signification des deux modalités peut être commune. Sekuler, R., Sekuler, A. et Lau (1997) ont par exemple montré que l'ajout d'un son au moment où deux cibles visuelles se croisent

(deux cercles) conduit à la perception d'un mouvement de rebond. Lorsque le son est absent lors du croisement, la détection du rebond chute considérablement et les observateurs perçoivent majoritairement un glissement (un des deux cercles passe sous l'autre). Cette interaction peut être testée à partir du site : http://hompi.sogang.ac.kr/mkyang/O/bach/ot/mot_bounce/index.html.

L'interaction audio-visuelle (c.-à-d. la modification de la perception de la modalité visuelle par la modalité auditive) à l'origine de cette illusion peut s'expliquer par la connaissance des observateurs du monde physique (facteurs conceptuels) où une collision entre deux objets est le plus souvent associée à un son. L'association entre le son et la proximité spatiale des objets conduirait alors à l'illusion. Les connaissances antérieures de l'observateur sont donc, dans le cadre de cette illusion, à l'origine du phénomène de fusion.

Pour que la fusion perceptive se produise, le pairing serait un pré-requis bien que celui-ci puisse survenir sans qu'il y ait pour autant de fusion. Ces deux notions coexistent si bien que la frontière les différenciant est très mince. Bien que leur distinction ait été discutée (Bertelson et Radeau, 1981), seule la notion de fusion sera abordée dans ce document pour signifier la perception d'un évènement audiovisuel unique.

Un exemple connu de la conséquence de fusion perceptive est l'*effet McGurk* (McGurk et McDonald, 1976) où la fusion des informations auditives « ba » et visuelles « ga » donne naissance à un nouveau percept : « da ». Ce percept, pourtant illusoire, est particulièrement robuste puisqu'il persiste même lorsque la nature des informations unimodales est connue. L'effet McGurk, tout comme l'illusion de rebond, peut être expliqué par l'influence des connaissances antérieures de l'observateur. En effet, en conditions habituelles, les informations auditives et visuelles concordent. Ainsi, la création du percept pourrait correspondre à la tentative de notre système perceptif de maintenir cette cohérence. Selon Fort (2002), la perception construite serait expliquée par l'interprétation la plus juste établie en fonction des proximités perceptives les plus adéquates pour chacune des modalités (/ba/ acoustiquement plus proche de /da/ que de /ga/, et /ga/ visuellement plus proche de /da/ que de /ba/). L'effet McGurk peut être testé à partir du site : http://www.faculty.ucr.edu/~rosenblu/VSMcGurk.html#

Un second exemple, également très connu et régulièrement expérimenté est l'*effet du ventriloque* (Howard et Templeton, 1966). Ce biais perceptif fait référence à l'illusion par laquelle un marionnettiste, en minimisant le mouvement de ses lèvres, trompe la perception de l'observateur. En effet, ce dernier identifie la source auditive comme provenant, non pas de l'illusionniste, mais de la marionnette dont les mouvements labiaux sont alors la seule information visuelle logique et disponible. Comme cela a été précédemment indiqué, la modalité visuelle serait dominante dans les tâches de localisation, c'est-à-dire qu'elle aurait tendance, en cas de présentation conflictuelle spatiale, à attirer dans sa direction la modalité auditive. Ce biais est mis à profit dans les systèmes artificiels de restitution audiovisuelle tels que le cinéma, la télévision ou encore la visioconférence. Ainsi, l'illusion du ventriloque est extrêmement familière puisqu'elle est expérimentée à chaque fois qu'un individu devient

spectateur avec l'illusion que le son provient de la bouche des acteurs plutôt que des haut-parleurs (voir Radeau, 1994).

L'effet du ventriloque se résume donc comme l'influence d'une source visuelle sur le jugement de localisation d'une source sonore associée (Driver, 1996). Il est plus généralement utilisé pour désigner, par extension, toute situation où la localisation d'une information auditive est influencée par une source visuelle (Radeau et Bertelson, 1977, 1978, Bertelson et Radeau, 1981). Chateau (1997) considère cependant que l'effet du ventriloque correspond à l'ensemble des situations où les observateurs ressentent une fusion perceptive entre le son et l'image restitués.

Comme présenté précédemment, notre système perceptif tente de maintenir, malgré certaines disparités, un percept unique et cohérent pour pouvoir extraire du sens de l'environnement. Le maintien de l'unité du percept donne parfois lieu, en raison des interactions entre les modalités auditives et visuelles, à des illusions perceptives. Certaines d'entre elles peuvent même être expérimentées quotidiennement.

Cependant, lors de trop grandes disparités spatiales ou temporelles, le phénomène de fusion peut s'altérer ou ne plus avoir lieu. Les bornes au-delà desquelles le groupement multimodal n'est plus garanti sont présentées dans les paragraphes suivants.

2.1.2.2. BORNES SPATIALES

Dans la vie de tous les jours, les informations visuelles et auditives d'un même évènement proviennent du même endroit (localisation identique). L'effet du ventriloque illustre l'attraction du son en direction de l'image pour permettre le groupement bimodal. Cependant, ce biais diminue voire disparaît à mesure que les sources unimodales s'éloignent (Bermant et Welch, 1976 ; Bertelson et Radeau, 1981). En d'autres termes, une disparité spatiale trop importante entre les sources auditives et visuelles altère ou empêche la fusion multimodale. Thurlow et Jack (1973) et Jack et Thurlow (1973) ont montré que l'effet du ventriloque disparaît à partir d'un écart angulaire entre l'image et le son (plan horizontal) de 30°/40° pour un environnement réverbérant. Par ailleurs, Komiyama (1989) a montré que, dans le cadre d'une présentation TV, un écart angulaire horizontal de 45° entre le son et l'image était perçu comme gênant. Une trop grande disparité spatiale a également pour conséquence une diminution du niveau de qualité audiovisuelle perçue.

L'altération du percept audiovisuel en réaction à une disparité spatiale trop importante est donc essentiellement liée à un écart trop élevé entre les dispositifs de restitution audio et vidéo. Ces disparités sont donc davantage le résultat d'erreurs d'installation que d'erreurs de captation ou de transmission du signal. C'est pourquoi les disparités spatiales ne seront pas plus amplement étudiées dans ce document.

2.1.2.3. BORNES TEMPORELLES

Tout comme la localisation, la synchronie est un principe essentiel pour la perception d'un évènement audiovisuel global. En effet, Jack et Thurlow (1973) ont constaté que la fusion

audiovisuelle (AV) est altérée voire perdue à partir d'un certain niveau de disparités temporelles. Les auteurs postulent que la diminution du degré de synchronisme affaiblirait la capture du son par l'image. Jack et Thurlow ont testé la durée de fusion perçue lors de différents niveaux de décalages temporels entre le son et l'image (désynchronisation) lors de la présentation de séquences audiovisuelles (lecteurs lisant de la prose) d'une durée de cinq minutes. Les participants devaient appuyer sur un bouton tant qu'une fusion était perçue entre l'audio et la vidéo, autrement, le bouton devait être relâché. Lorsque le bouton était enfoncé, une minuterie était déclenchée pour enregistrer l'intervalle de temps durant lequel un son était perçu comme venant de la même direction que l'image (fusion). La durée de fusion AV perçue a été étudiée pour trois niveaux de désynchronisation : 100 ms, 200 ms et 300 ms de retard du son par rapport à l'image. Les résultats ont indiqué que la désynchronisation est relativement bien tolérée pour un retard du son de 100 ms (100% des participants ont perçu une fusion pendant une durée moyenne de 220,4 s), significativement dégradée lors d'un retard de 200 ms (70% pendant 113,8 s) et quasiment inexistante pour un délai de 300 ms (30% pendant 21,3 s). Ainsi, la fusion AV perçue est sensible au décalage temporel entre le son et l'image.

2.1.3. PERCEPTION DE LA DESYNCHRONISATION

La diminution de la coïncidence temporelle entre les informations auditives et visuelles peut donc altérer le percept audiovisuel en tant qu'évènement multimodal. Dans les environnements artificiels de restitution AV (télévision, *etc.*), la présence de désynchronisation entre l'image et le son peut avoir un effet préjudiciable sur la qualité perçue (Rihs, 1995). Par conséquent, il est nécessaire de contrôler la relation temporelle entre les signaux audio et vidéo de façon à ce que la qualité perçue par l'utilisateur ne soit pas altérée. Les disparités temporelles sont d'autant plus importantes à considérer qu'elles représentent la principale cause de dégradation audiovisuelle consécutive à des erreurs survenant dans la chaîne de transmission.

2.1.3.1. ASYMETRIE

Un aspect fondamental à prendre en compte dans la gestion de la relation temporelle entre image et son est la sensibilité asymétrique de la perception de la désynchronisation. En effet, un son en retard par rapport à l'image est mieux toléré qu'un son en avance. Cavé, Ragot et Fano (1992) ont mesuré l'influence d'un décalage temporel entre l'image et le son sur le phénomène de fusion perceptive. Des observateurs, situés à sept mètres d'un écran de cinéma, avaient pour tâche de juger la synchronie entre un stimulus visuel (clap de cinéma) et un stimulus auditif (bruit de ce même clap). Ce stimulus AV a été choisi pour son caractère impulsif et non cyclique. Des décalages image/son compris entre -260 ms (son en avance par rapport à l'image) et +100 ms (son en retard) ont été introduits par pas de 40 ms. Les observateurs devaient juger le son comme étant en avance, en retard ou synchrone par rapport à l'image. La synchronie a été détectée lorsque le son présentait un retard de 40 ms. Il est à

noter que ce retard correspond au temps d'intégration du signal visuel, des récepteurs sensoriels jusqu'à leur intégration par les neurones multimodaux du colliculus supérieur.

Ainsi, dans le cadre de restitution AV, un léger retard du son sur l'image est perçu comme étant plus synchrone qu'un son ne présentant pas de décalage avec l'image (synchronisation physique). Dans ce dernier cas, le son était jugé comme légèrement en avance. De manière générale, le point de synchronie subjective se situerait entre 30 et 50 ms (voir Kohlrausch et van de Par, 2005).

Les résultats de Cavé *et al.* (1992) ont également indiqué que la zone de synchronie subjective s'étend d'environ +20 ms à +100 ms de retard du son. Les observateurs seraient donc beaucoup plus tolérants lorsqu'un son est présenté en retard par rapport à l'image que la situation inverse. A l'opposé, un son en avance est moins bien toléré (Cavé *et al.*, 1992 ; Dixon et Spitz, 1980 ; Hollier et Rimmel, 1998). Cette asymétrie est généralement expliquée par l'impossibilité physique de percevoir un son avant l'image de cette même source en raison des différences de vitesses de propagation des ondes acoustiques et lumineuses.

2.1.3.2. INFLUENCE DU CONTENU

L'amplitude de la zone de synchronie subjective est fortement dépendante de la méthode employée (choix de catégories : son en avance/synchrone/image en avance ; perception de la désynchronisation AV : synchrone ou non synchrone ; choix de la modalité en avance : audio ou vidéo, voir Van de Par, Kohlrausch et Juola, 2002 pour plus de détails sur l'influence de la méthode), des consignes (localisation du son ou fusion audiovisuelle ressentie par exemple, voir Jack et Thurlow, 1973) ou encore du type de contenu. Par exemple, Dixon et Spitz (1980) ont mesuré les seuils de détection de la désynchronisation pour des séquences AV verbales (locuteur lisant de la prose) et non verbales (marteau frappant une cheville). Les seuils de détection obtenus étaient compris entre -75 ms (son en avance) et +188 ms (son en retard) pour la scène du marteau et entre -131 ms et +258 ms pour la scène du locuteur. Ces résultats indiquent un effet de la nature du contenu. Les observateurs ont en effet toléré une plus grande amplitude de désynchronisation lors de la séquence verbale (paroles) par rapport à la séquence non verbale (bruits impulsifs). Une explication consiste à croire que la désynchronisation serait beaucoup plus difficile à détecter en présence de sons de parole caractérisés comme semi-périodiques par rapport à un bruit impulsif. Cependant, les valeurs de détection obtenues dans cette étude pourraient ne pas être le reflet exact de la sensibilité des individus à la désynchronisation AV en raison d'un certain nombre de facteurs méthodologiques (Vatakis et Spence, 2005). Par exemple, les stimuli auditifs étaient présentés au moyen d'un casque audio tandis que les stimuli visuels étaient présentés sur un moniteur situé en face du participant. Or, l'intégration des stimuli multisensoriels est facilitée par leur coïncidence spatiale. Cette étude peut ainsi avoir surestimé la sensibilité des individus à la désynchronisation en fournissant des repères spatiaux et temporels qui ne sont pas présents lorsque les stimuli auditifs et visuels proviennent du même endroit, comme au cours d'un dialogue par exemple.

Selon Hollier et Rimmel (1998) la désynchronisation serait plus rapidement perçue lors de séquences verbales que non verbales lorsque le son est présenté en avance. Plus précisément, ces auteurs ont étudié le nombre d'erreurs de synchronisation détectées pour trois différents types de séquences : une séquence courte non verbale présentant un objet au rebond (stylo), une séquence longue non verbale présentant des coups de hache et une séquence verbale (buste du locuteur et sons de parole). Les durées des séquences variaient entre 500 ms et 4,5 s. Neuf niveaux de désynchronisation, entre -150 ms (son en avance) et + 300 ms (son en retard) échelonnés par pas de 50 ms, ont été testés. Après la visualisation de chaque séquence, les participants devaient répondre par « oui » si une erreur de synchronisation avait été perçue et par « non », dans le cas contraire. Comme le présente la Figure 2.2 ci-après, un plateau de non perceptibilité des erreurs de synchronie s'étendrait d'environ -40 ms (son en avance) à environ + 120 ms, tous types de contenus confondus.

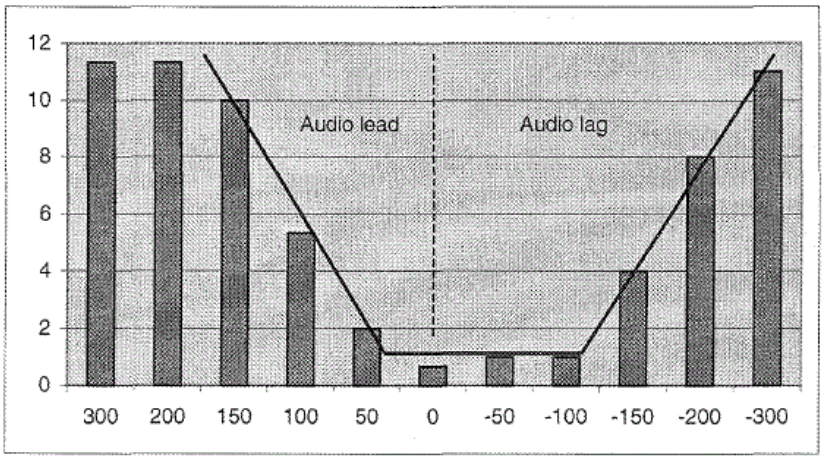


Fig. 2.2. Plateau de non perceptibilité des erreurs de synchronie. La Figure présente le nombre de participants ayant détecté des erreurs de synchronisation, moyenné pour l'ensemble des contenus, pour les neuf niveaux de désynchronisation étudiés. Ici les valeurs négatives correspondent à un son en retard par rapport à l'image (Hollier et Rimmel, 1998).

Contrairement aux travaux de Dixon et Spitz (1980), les participants de cette étude ont montré une plus grande sensibilité à la désynchronisation pour les stimuli AV verbaux par rapport aux stimuli non verbaux (sons impulsifs). Toutefois, ce constat est vrai lorsque, et uniquement dans cette situation, le son était présenté en avance. Selon les auteurs, la courte durée de certaines unités sémantiques (attaques des consonnes par exemple) rendrait la détection plus facile. Les proportions de détection ne différaient pas de manière significative entre les trois séquences (verbales et non verbales) lors d'un son en retard. Les participants de cette étude ont également détecté plus d'erreurs de synchronie pour le contenu « hache » que le contenu « stylo ». L'ensemble de ces résultats met en avant l'influence importante du type de contenu sur la détection de la désynchronisation.

Plus récemment, Vatakis et Spence (2005) ont étudié la sensibilité des individus à la désynchronisation audiovisuelle lors de la présentation de séquences verbales (visage d'un locuteur, phonèmes et syllabes brèves) et musicales (notes de guitare ou de piano) de durées inférieures à trois secondes. Neuf niveaux de désynchronisation, compris entre - 400 ms

secondes (son en avance) et + 400 ms (son en retard) par pas de 100 ms, ont été étudiés. Les participants devaient indiquer après chaque séquence quelle modalité, audio ou vidéo, avait selon eux été présentée en premier. Le seuil différentiel moyen (en-dessous duquel la désynchronisation n'était plus correctement détectée) a été calculé pour chaque séquence de test. Comme le présente la Figure 2.3, les participants ont obtenu de meilleures performances de détection lors de séquence verbale par rapport aux séquences musicales. Ce résultat suggère que les individus sont plus sensibles à la désynchronisation AV survenant sur des séquences verbales plutôt que musicales.

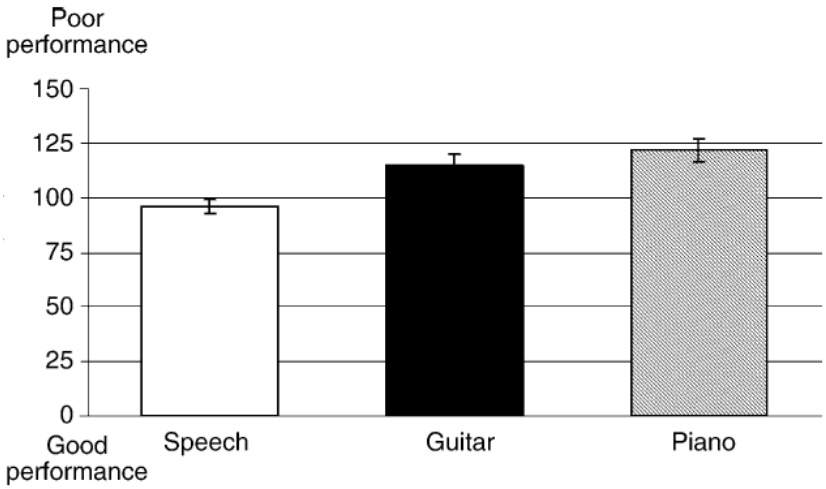


Fig. 2.3. Seuil différentiel moyen (ms) obtenu pour chaque séquences verbales (Speech) et non verbales (Guitar et Piano). Les barres d'erreur représentent les écarts-types de la moyenne.

En résumé, la fusion audiovisuelle serait perdue à partir d'un écart angulaire de 30° à 40° degré entre le son et l'image et d'un délai de 300 ms du son par rapport à l'image. Par ailleurs, la présence de désynchronisation est moins bien tolérée lorsque le son est en avance sur l'image par rapport à la situation inverse. Il semblerait que les participants soient plus sensibles à la désynchronisation lors de situations non écologiques (impossibilité physique qu'un son soit perçu avant l'image issue d'une même source). Ce constat tend à être confirmé par la détection plus rapide de la désynchronisation lors d'un son de parole en avance par rapport à l'image correspondante. De manière générale, une désynchronisation image/son lors de séquences verbales est moins bien tolérée que lors de séquences non verbales (bruit ou musique). La perception de la désynchronisation est, en effet, dépendante du type de contenu testé. Un aspect supplémentaire à considérer dans le cadre de restitution AV est l'existence d'un point de synchronie subjective constaté, non pas lors d'un délai nul entre le son et l'image, mais lorsque que le son est présenté avec un retard compris entre 30 et 50 ms.

Ces différents résultats (sensibilité asymétrique, influence du contenu, *etc.*) sont autant d'aspects critiques à considérer dans le cadre de système de diffusion AV.

2.1.3.3. DESYNCHRONISATION ET QUALITE PERÇUE

La gestion de la synchronisation temporelle des signaux audio et vidéo est une préoccupation centrale dans le cadre de diffusion de contenus audiovisuels. En effet, plusieurs études ont montré qu'une différence perceptible entre les temps de transmission des composantes du son et de l'image d'un signal AV est gênante pour l'utilisateur. Il a été trouvé que, pour un contexte télévisuel (journal TV), la désynchronisation est perceptible à partir de -45 ms (son en avance) et de +125 ms (son en retard) et inacceptable à partir de - 90 ms et de +185 ms (Rihs, 1995 ; UIT-R BT.1359-1, 1998). Une illustration des plateaux de perceptibilité et d'acceptabilité est apportée par la Figure 2.4 ci-après. Par ailleurs, la qualité perçue se dégrade rapidement lorsque la désynchronisation augmente (Rihs, 1995). Précisément, la désynchronisation serait perçue comme étant gênante à partir de 150 ms d'avance du son sur l'image (Rihs, 1995 ; UIT- R SG11, 1995).

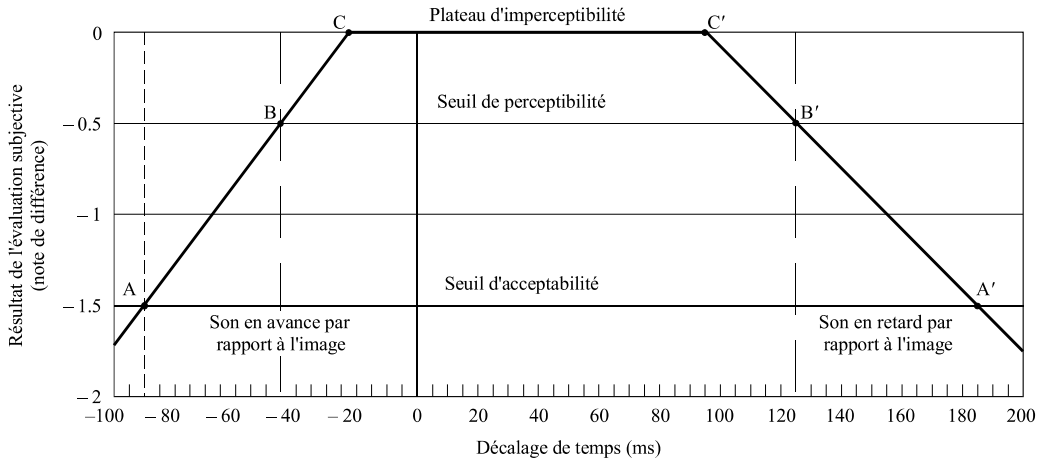


Fig. 2.4. Plateaux de perceptibilité et d'acceptabilité (extrait de la norme UIT-R BT.1359-1, 1998), un son en avance est représenté par un le signe « - ».

La question de la synchronisation des signaux audio et vidéo est donc un aspect critique pour garantir la qualité du signal audiovisuel telle que perçue par l'utilisateur. Dans l'objectif de ne pas introduire de dégradations de désynchronisation, la norme UIT-T J.100 (UIT, 1990), recommandée par l'UIT, fixe le délai entre les signaux audio et vidéo, dans le cadre de diffusion audiovisuelle, à 20 ms d'avance du son et 40 ms de retard du son. Ces délais se situent bien en-dessous des seuils de détection de la désynchronisation présentés dans le paragraphe précédent (-45 ms et +125 ms). Le respect de la marge recommandée par la norme UIT J.100, valable pour tous types de contenus, permet de garantir que les erreurs de synchronisation ne dégradent pas la qualité perçue en demeurant imperceptibles pour l'utilisateur. L'ensemble des résultats précédemment présenté est récapitulé par le Tableau 2.1 ci-après.

Tableau 2.1. Synthèse des résultats précédemment présentés. Les signes - et + représentent respectivement un son en avance et un son en retard par rapport à l'image.

Auteurs	Stimuli	Etudié	Résultats
JACK et THURLOW	Parole (locuteur)	Durée de fusion	220 s/délai +100 ms 114 s /délai +200 ms 21 s /délai +300 ms
CAVÉ *et al.*	Clap cinéma (Clap)	Zone de synchronie	[+20 ; +100 ms]
DIXON et SPITZ	Prose (locuteur) Marteau (marteau)	Détection synchronie	[-131 ; + 258 ms] [-75 ; + 188 ms]
HOLLIER et RIMMEL	Rebond (stylo) Hache (hache) Parole (locuteur)	Détection erreur de synchronie	[-40 ; + 120 ms] son en avance : verbal > non verbal
VATAKIS et SPENCE	Syllabes (locuteur) Notes (instrument)	Performance de détection	Verbal > musical
RIHS **UIT-R BT.1359**	Journal TV	Détection synchronie Acceptabilité	[-45 ; + 125 ms] [-90 ; + 185 ms]

La présence de disparités spatiales ou temporelles peut donc dégrader la qualité perçue de la restitution AV. En plus de ces dégradations, le niveau de qualité du signal d'une modalité, audio ou vidéo, va influencer la perception de qualité de l'autre modalité. Ces influences mutuelles entre qualité audio ou vidéo conditionnent la perception de la qualité audiovisuelle globale.

2.2. QUALITE AUDIO ET VIDEO ET PERCEPTION DE QUALITE AUDIOVISUELLE

Cette section aborde la question de la qualité perçue des signaux audio et/ou vidéo et notamment celle de leurs influences mutuelles. Dans le cadre de l'optimisation de services de diffusion AV, il est nécessaire de connaitre la manière dont ces influences (entre modalités audio et vidéo) impactent la perception d'une part, de la qualité de chaque modalité et d'autre part, de la qualité audiovisuelle globale (QAV).

2.2.1. PERCEPTION DE LA QUALITE AUDIOVISUELLE GLOBALE : CONTRIBUTIONS DES QUALITES AUDIO ET VIDEO ET INTERACTIONS

La qualité perçue audiovisuelle (QAV) ne résulte pas de la simple addition des qualités audio (QA) et vidéo (QV) mais de l'interaction des niveaux de qualité de chaque modalité. Plusieurs études ont en effet mis en évidence les influences mutuelles qu'exercent l'un sur l'autre les niveaux de qualité audio et vidéo (Beerends et de Caluwe, 1999 ; UIT-T COM 12-19-E, 1997). Les études présentées ci-après portent sur les interactions des qualités audio et vidéo, étudiées à travers les notes MOS, ainsi que leur contribution respective à la perception de la qualité audiovisuelle globale.

Une expérience a été conduite (dans le cadre du projet européen MOSAIC, voir Kohlrausch et van de Par, 2005) afin de comparer les jugements de qualité obtenus pour des

images présentées avec ou sans son. Les stimuli consistaient en trente séquences de cinquante secondes chacune et présentées avec trois niveaux différents de qualité vidéo : source (sans dégradations), débit réduit à 4 Mbps et débit réduit à 2 Mbps. Le son n'était jamais dégradé. Les participants devaient juger la qualité vidéo sur une échelle en 10 points. Un premier groupe visualisait la vidéo sans le son, un deuxième groupe visualisait les séquences avec le son. Les résultats ont montré que toutes les séquences ont été jugées avec un score plus élevé de qualité en présence du son par rapport à la condition vidéo seule. Les auteurs concluent que la capacité des participants à détecter des dégradations vidéo est plus faible lorsque qu'un son de bonne qualité l'accompagne. Ce résultat illustre notamment la manière dont la qualité d'une modalité peut influencer la perception de la qualité de l'autre modalité lors d'une présentation audiovisuelle.

Hollier et Voelcker (1997) ont montré que la qualité de séquences audio dégradées présentées seules était notée avec des notes plus sévères par rapport à la présentation de ces mêmes séquences en présence de vidéo. Plus précisément, ces auteurs ont étudié l'influence de la qualité objective de la vidéo sur la perception de la qualité audio et audiovisuelle. Les séquences utilisées consistaient en huit clips vidéo de réalité virtuelle (survol de bâtiments), d'une durée de dix secondes, en format TV numérique. Des commentaires, énoncés par une voix d'homme ou de femme, accompagnaient la vidéo. Les participants devaient évaluer QAV ou QA (lors de sessions distinctes) après chaque séquence présentée (méthode ACR, échelle catégorielle en 5 points). Différents niveaux de qualité audio et vidéo étaient présentés. Les résultats obtenus ont également montré que la qualité objective de la modalité vidéo jouait un rôle prépondérant dans l'évaluation de la qualité audiovisuelle tandis que la qualité objective de la modalité audio tenait une place secondaire. Par ailleurs, la qualité perçue de l'audio diminuait lorsque la vidéo se dégradait (la situation inverse n'a pas été étudiée dans cette expérience). Selon les auteurs, ces résultats doivent toutefois être mis en regard d'une possible influence du type de contenu sur les notes MOS obtenues. En effet, certaines séquences pouvaient contenir plus ou moins d'artefacts vidéo (liés au processus de traitement numérique). La nature, masculine ou féminine, des commentaires aurait également pu participer aux effets observés.

En 1999, Beerends et de Caluwe ont investigué les influences mutuelles des qualités audio et vidéo sur la perception de la qualité de chacune de ces modalités et du signal audiovisuel global. Pour cela, des dégradations survenant aussi bien sur la modalité vidéo qu'audio ont été introduites. Plus précisément, deux séquences audiovisuelles (clips publicitaires) dégradées (combinaison de quatre dégradations audio et vidéo) de vingt-cinq secondes étaient visualisées/entendues. Les signaux audio (A) et vidéo (V) de ces mêmes séquences étaient également présentés séparément. Après chaque présentation (AV, V ou A), les participants devaient juger soit la qualité audio (pour le signal audio seul - MOSAA - ou le signal AV - $MOSAV_A$-), vidéo (signal vidéo seul – MOSVV - ou AV -$MOSAV_V$-) ou audiovisuelle ($MOSAV_{AV}$). Pour l'évaluation, l'échelle catégorielle en 9 points de la méthode ACR a été utilisée.

Les notes MOS obtenues ont montré que $MOSAV_A$ était influencée par le niveau de qualité vidéo et *vice versa*. Par exemple, une bonne qualité vidéo (ou audio) entraînait une amélioration de la note de qualité audio (ou vidéo). Ce résultat a été obtenu par la soustraction de la note MOSAA à la $MOSAV_A$ pour refléter la contribution de la vidéo à la note de qualité audio. Ainsi, le jugement de la qualité d'une modalité est influencé par le niveau de qualité de l'autre modalité. Ce constat confirme les résultats présentés par Kohlrausch et van de Par (2005) et Hollier et Voelcker (1997). Cependant, la contribution de QV sur l'évaluation de QA est plus importante (variation jusqu'à environ un point de MOSA) que la situation inverse (variation jusqu'à 0,2 point de MOSV). Beerends et de Caluwe rapportent également que la participation de QV à l'évaluation de QAV est plus importante que la contribution de la qualité audio. Les auteurs concluent que QV domine la qualité AV perçue globale dans le cadre d'applications non conversationnelles.

La « Commission 12 » de l'UIT (COM12-61-E, 1998) a étudié l'influence du contexte (passif ou interactif) sur la perception de QV, QA et QAV. Pour le contexte passif, les participants visualisaient deux extraits de visioconférence de dix secondes chacun (interlocuteurs homme ou femme) tandis que dans le second, ils devaient participer à une activité de conversation (mots à deviner). Au total, seize conditions de qualité (combinaison de quatre niveaux de qualité audio et vidéo) étaient présentées pour les deux contextes. La méthode d'évaluation proposée pour le contexte passif était identique à celle utilisée par Beerends et de Caluwe (1999). Dans le cas du contexte interactif, les participants évaluaient les qualités AV, V et A. Les résultats ont montré que le niveau perçu de qualité vidéo était indépendant du contexte (forte influence des différents niveaux de qualité vidéo sur MOSV et MOSAV). Ce résultat n'est pas retrouvé pour la qualité audio perçue : les notes de qualité audio ne reflétaient pas les dégradations audio lors du contexte interactif (faible influence des différents niveaux de qualité audio sur MOSA et MOSAV). Les auteurs postulent que dans un contexte conversationnel, la vidéo serait bien jugée à partir de critères de qualité (netteté, fluidité) mais l'audio serait évalué sur la base de critères d'acceptabilité (intelligibilité, volume, écho). Or, les auteurs précisent que les niveaux de qualité audio proposés pouvaient être considérés comme acceptables dans le sens où ils se situaient toujours au-dessus du seuil d'intelligibilité (ne gênaient pas la compréhension). Par conséquent, dans la mesure où l'acceptabilité n'était pas diminuée par les niveaux de qualité audio présentés, ceux-ci n'étaient pas ou peu reflétés par les notes reportées sur l'échelle de qualité (MOSA). Ainsi, dans un contexte passif, les participants jugeraient la qualité audio et vidéo, tandis qu'en contexte interactif, ils jugeraient la qualité vidéo et l'acceptabilité audio.

Par ailleurs, la perception de QV était indépendante de QA tandis que cette dernière a été influencée par QV et ce, pour les deux contextes testés. Cette influence était cependant beaucoup moins importante en contexte passif.

Les résultats ont aussi indiqué que les contributions de QV et QA à QAV étaient dépendantes du contexte. La contribution de la vidéo était élevée dans les deux cas tandis que la contribution de l'audio était faible en contexte passif et inexistante en contexte conversationnel. Ainsi, QV a plus fortement contribué à la note de QAV notamment en contexte conversationnel. Tout comme Beerends et de Caluwe, les auteurs de cette étude

concluent à une prédominance de la qualité objective vidéo à la perception de la qualité audiovisuelle globale.

Contrairement aux précédentes conclusions, Hands (2004) affirme que la qualité audio serait dominante pour un contexte passif (présentation d'une vidéo de locuteur *tête-épaule*). Selon cet auteur, l'évaluation de la qualité serait en fait fortement dépendante du contenu de test : la qualité audio serait dominante lors de séquence AV faiblement dynamiques (contenu locuteur *tête-épaule*), tandis que la qualité vidéo serait dominante lors de séquences AV fortement dynamiques (nombreux mouvement, changement de plans, *etc*.). Hands a en effet constaté une prédominance de l'audio sur QAV lors de l'évaluation de séquences audiovisuelles de cinq secondes (locuteur tête-épaule) à partir d'une échelle continue à 5 niveaux d'excellent à mauvais (méthode DSCQS : *Double-Stimulus Continuous Quality Scale*, recommandée pour évaluer la qualité vidéo perçue de paires de séquences). Le désaccord entre ce résultat et ceux de Beerends et de Caluwe (1999) serait alors expliqué par un effet de contenu. Selon Hands, les clips commerciaux utilisés par Beerends et de Caluwe contiendraient de nombreuses informations visuelles (dynamique élevée : nombreux mouvements, changements de plans, *etc*.) qui auraient tendance à capter l'attention du spectateur. A l'inverse, la vidéo d'un locuteur serait caractérisée par une certaine pauvreté informationnelle du média visuel et orienterait plutôt l'attention du spectateur vers la modalité audio.

Hands confirme d'ailleurs ce postulat par une seconde expérience présentant deux types de contenus vidéo de cinq secondes. Le premier est défini par une dynamique faible (locuteur *tête-épaule*) et le second par une dynamique élevée (course de cyclistes + commentaires). Les qualités A, V et AV ont été évaluées au moyen de la méthode SSQS (*Single Stimulus Quality Scale* recommandée pour évaluer la qualité vidéo perçue d'une seule image ou séquence d'images à partir d'une échelle de catégories à cinq niveaux d'excellent à mauvais). Les résultats obtenus pour les séquences *tête-épaule* confirment l'influence prédominante de QA tandis que ceux obtenus pour les séquences à dynamique élevée indiquent une plus grande influence de QV sur la qualité audiovisuelle globale. Le focus attentionnel serait alors porté vers la modalité véhiculant l'information primordiale et sans laquelle l'accès au sens du contenu serait plus difficile ou impossible. Cette hypothèse s'apparente à celle de la modalité pertinente décrite dans le paragraphe 2.1.2.1 ci-avant (Welch *et al.*, 1986) à la différence qu'au lieu de prioriser la modalité permettant de décrire au mieux les stimuli en matière de codage de bas-niveaux (primitives visuelles – contour, forme, *etc*.- ou auditives – hauteur, localisation, *etc.*), la priorité sera attribuée au codage de plus haut-niveau (modalité porteuse de l'information primordiale : compréhension). Ce constat laisse supposer un effet de la modalité dominante sur la perception de la qualité audiovisuelle globale.

Hands a également montré que l'impact de QV sur QA diminue à mesure que le niveau de qualité audio diminue. Le même effet est constaté pour l'influence de QA sur la perception de QV. Il est possible que cet effet soit un effet *plancher* : lorsque les qualités sont trop basses, les variations des scores sont alors trop faibles pour pouvoir étudier les influences mutuelles des qualités audio et vidéo. Pour plus de clarté, les différents résultats présentés ci-dessus sont synthétisés par le Tableau 2.2 ci-dessous.

Tableau 2.2. Synthèse des résultats. Les influences mutuelles entre qualité audio (A) et vidéo (V) sont présentées : V(A) correspond à une influence dominante de V sur A, l'inverse est indiqué par A(V). La prédominance de la qualité A ou V sur QAV est également reportée.

AUTEURS	VIDEO	AUDIO	DUREE	METHODE	INFLUENCE A-V	PREDO-MINANCE A-V sur AV
MOSAIC	Contes	Voix/Ø	50 s	Echelle 10-points	A améliore V	-
HOLLIER 97	Réalité virtuelle	Commentaire	~10 s	ACR Echelle 5-points	V(A)	V
BEERENDS	Clips publicitaires	Voix	25 s	ACR Echelle 9-points	V(A)	V
COM 12-61 E	Locuteur Partenaire	Voix	10 s	ACR Echelle 5-points	V(A) V(A)	V
HANDS	Locuteur	Voix	5 s	DSCQS SSCQS Echelle 5-points	A(V)	A
	Course cycliste	Commentaire	5 s	DSCQS Echelle 5-points	V(A)	V

2.2.2. CAS SPECIFIQUE DE L'INTELLIGIBILITE

En situation naturelle, les informations visuelles facilitent ou améliorent l'intelligibilité des sons de parole notamment en environnement bruyant (effet cocktail party, Cherry, 1953 ; Sumby et Pollack, 1954). Summerfield (1979) a montré que, dans le cadre d'une présentation audiovisuelle en environnement bruité, l'intelligibilité est améliorée par la vision du visage entier du locuteur (gain de 42,6% par rapport à la condition où l'audio est présenté seul) ou de sa bouche (gain de 27,3%). Ainsi, les observateurs s'appuient sur des indices faciaux pour améliorer leur perception de l'information auditive en contexte bruité. Plus ces indices sont nombreux (bouche *vs.* visage) plus l'intelligibilité du discours sera améliorée.

Cependant, Grant et Braida (1991) ont montré que le gain apporté par la modalité visuelle est élevé (35 à 50%) lorsque que la bande passante audio est réduite, mais plafonne autour de 20 à 30% quand la qualité du signal audio est bonne. La plus-value apportée par la modalité vidéo serait d'autant plus forte que le signal audio est dégradé et diminuerait dans le cas où le média audio présente une bonne qualité. Ce constat n'est pas sans rappeler le principe d'efficacité inversée pour l'intégration neuronale multimodale (sect. 2.1.1 ci-avant) à savoir qu'aucun gain n'est obtenu par l'ajout d'une modalité lorsqu'un stimulus uni-modal est suffisamment efficace seul.

L'intelligibilité d'une séquence diminue également lorsque l'audio et la vidéo ne sont pas synchronisés (Campbell et Dodd, 1980 ; Massaro, Cohen et Smeele, 1996). Campbell et Dodd (1980) ont étudié le bénéfice d'une présentation audiovisuelle sur l'intelligibilité de pseudo-mots (mots sans signification) en présence de désynchronisation. Les pseudo-mots étaient formés de la manière suivante : consonne-voyelle-consonne et présentés en présence de bruit blanc (bruit dont toutes les fréquences audio sont de la même intensité). Les participants devaient répéter les mots entendus sur une liste de 20 mots. Ces auteurs ont montré que l'intelligibilité se dégrade en cas de désynchronisation mais que le bénéfice par rapport à une

présentation uni-modale perdure même lorsque le décalage entre l'audio et la vidéo atteint + 1600 ms (retard du son).

La modalité visuelle améliore donc l'intelligibilité de la modalité auditive notamment lorsque cette dernière est dégradée au point de gêner la compréhension si l'audio avait été présenté seul. Ces études mettent en avant l'influence de la qualité d'une modalité (vidéo de bonne qualité) sur l'autre modalité (qualité audio dégradée).

Les précédents paragraphes conduisent à différents constats. Tout d'abord, les niveaux de qualité audio et vidéo s'influencent mutuellement et la qualité audiovisuelle perçue résulte de ces diverses influences. Notamment, il semblerait que la qualité audiovisuelle perçue dépende plus fortement de la qualité objective vidéo qu'audio. Toutefois, la contribution de chaque modalité à la qualité audiovisuelle globale serait dépendante du contenu et notamment de la modalité dominante. En effet, si la richesse informationnelle de la modalité audio (A) est plus importante (porteuse de l'information primordiale) que celle de la modalité vidéo (V), alors la prédominance de V sera plus faible ou nulle, c'est-à-dire que l'influence prédominante viendra de A. Les méthodes subjectives d'évaluation de la qualité audiovisuelle doivent donc tenir compte de l'effet du contenu ainsi que de l'ensemble des facteurs, autres que les seuls niveaux de qualité audio et vidéo, capable d'influencer son évaluation. Les paragraphes suivants décriront ces différences influences.

2.2.3. PERCEPTION DE LA QUALITE AUDIOVISUELLE : AUTRES FACTEURS

Comme précédemment présenté, les caractéristiques des contenus de test telles que leur nature verbale ou non (parole, bruit, musique), leur dynamique ou encore leur modalité dominante ont un rôle primordial sur la perception audiovisuelle (fusion, perception de disparités) et sur la qualité audiovisuelle perçue. Au-delà de ces aspects, peu d'études ont investigué la relation entre la vidéo et l'audio comme l'ont constaté You, Reiter, Hannuksela, Gabbouj et Perkis (2010). Par exemple, il est probable qu'une dégradation survenant sur l'expression sonore d'un évènement visible à l'image (son *in*) n'ait pas la même influence sur la qualité audiovisuelle perçue que cette même dégradation survenant sur un son en-dehors de l'image (son *hors-champ*). L'influence de la relation entre son et image sur la perception de qualité pourrait être d'autant plus importante que la nature du signal audio soit verbale ou non verbale. L'influence du contenu sur le jugement de QAV pourrait également dépendre de facteurs subjectifs : un spectateur serait plus tolérant, quant au niveau de qualité, lorsque le contenu lui paraît attractif (Palhais *et al.*, 2012, sect. 1.6.3, chap. I).

La perception de la qualité peut également être influencée par des effets attentionnels. Selon Hollier, Rimell, Hands et Voelcker (1999) des phénomènes attentionnels pourraient empêcher la perception des dégradations (pourtant présentes sur les deux modalités) sur l'une ou l'autre modalité. Ces auteurs parlent d'un phénomène de masquage cross-modal où l'attention de l'individu serait attirée par les défauts présents sur une des deux modalités au point que les dégradations de l'autre modalité soient négligées. L'orientation de l'attention sur la qualité de l'une ou l'autre modalité peut être attribuée à différents facteurs : contenu

(modalité dominante par exemple), consignes, contexte d'évaluation, réalisation d'une tâche, *etc*.

L'influence de la tâche a été étudiée par Reiter et Weitzel (2007) et Reiter, Weitzel et Cao (2007). Ils ont montré que l'influence de la tâche était plus importante si elle était réalisée sur la modalité devant être jugée. Par ailleurs, la précision avec laquelle la qualité est évaluée dépendrait du degré de distraction (tâche) sur la même modalité. Ces mêmes auteurs ont également indiqué que les dégradations de la qualité audio d'une séquence donnée étaient moins perceptibles dans un contexte audiovisuel actif (condition de jeux vidéo) comparativement à un contexte passif (contexte TV). Au-delà d'un effet relatif à la tâche, ces résultats soulignent un effet potentiel de la division cross-modale de l'attention se manifestant par des stimuli audio dégradés mieux notés en situation d'attention partagée (jeu vidéo) qu'en situation d'attention focalisée (TV). Ainsi, le jugement de qualité en contexte passif de visualisation augmente la capacité des spectateurs à détecter les dégradations, leur attention étant centrée sur le ou les médias à évaluer. Ce constat est important car il met en évidence la nécessité de proposer une qualité audiovisuelle optimale en contexte passif de visualisation (notamment pour un contexte télévisuel).

Une influence des consignes sur la perception de qualité a été constatée par Rimell et Owen (2000). Le jugement de la qualité des signaux audio ou vidéo d'une séquence audiovisuelle varierait, selon ces auteurs, de manière significative selon que l'attention soit orientée vers la modalité à évaluer ou partagée entre les deux modalités. Ce constat rappelle celui de Ragot *et al.* (1988) pour lequel les consignes données aux participants (à savoir de porter leur attention sur l'une ou l'autre modalité) auraient une influence sur la prédominance d'une modalité sur l'autre (§ 2.1.2.1).

Enfin, des facteurs contextuels notamment liés à la notion de QoE (humeur, engagement financier, expériences antérieures, ergonomie) peuvent également agir sur la perception de qualité. Le Tableau 2.3 ci-dessous résume les différentes influences de la perception de qualité.

Tableau 2.3. Synthèse des différentes influences de la perception de la qualité regroupées par familles : contenu, attention et contexte.

FAMILLE D'EFFETS	EFFETS	CONSEQUENCES
CONTENU	Modalité Attractivité (intérêt) Relation AV	Influence la perception de dégradations
ATTENTION	Orientation de l'attention (partagé vs. dirigée) : tâche, consignes, contexte passif *vs*. actif	Masquage cross-modal (dégradation négligée pour l'une des deux modalités)
		Focalisation sur les dégradations de l'une ou l'autre modalité
CONTEXTE	Facteurs QoE	Influence la qualité perçue

2.3. VERS LA MESURE DE L'ACTIVITE PHYSIOLOGIQUE ET OCULAIRE

La perception de la qualité audiovisuelle relève de la manière dont les informations auditives et visuelles sont traitées et fusionnées par le système audiovisuel humain (SAVH). Le respect des principes de coïncidence spatiale et temporelle est indispensable pour la construction d'un évènement audiovisuel unique. La présence de disparités spatiales ou temporelles importantes peut en effet altérer voire empêcher le phénomène de fusion, auquel cas les informations issues de chaque sens vont être perçues comme des évènements distincts. Cependant, malgré ces disparités, le SAVH tente de maintenir, dans une certaine mesure, l'unité du percept. Cela explique notamment le maintien de la fusion des informations auditives et visuelles dans le cadre de système de restitution AV (cinéma, TV, visioconférence). Il est toutefois fondamental de se rapprocher le plus possible des principes de coïncidence pour ne pas dégrader la qualité perçue par l'utilisateur, par exemple, les marges acceptables de désynchronisation sont explicitement fixées par les instituts de normalisation.

Au niveau du signal intrinsèque audio ou vidéo, la perception de la qualité audio, vidéo et audiovisuelle dépend des influences mutuelles qu'entretiennent les niveaux de qualité de chaque signal (audio ou vidéo). La contribution de la qualité de l'audio et de la vidéo à la qualité globale va principalement dépendre du type de contenu (modalité dominante, intérêt, *etc.*) mais aussi de facteurs attentionnels et contextuels.

La qualité audiovisuelle perçue est donc un phénomène complexe qui se résume difficilement par un effet unique de la qualité objective du signal (quantité de dégradations présentes sur le signal) et par une évaluation sur une seule et unique échelle catégorielle de qualité. Comme indiqué dans le chapitre I, une méthode alternative ne reposant pas sur une approche unidimensionnelle de la qualité consisterait à mesurer l'activité physiologique et oculaire des spectateurs. La finalité d'une telle méthode est de pouvoir d'une part, contourner les biais des mesures subjectives actuelles et d'autre part, d'appréhender de manière plus fidèle l'influence de la qualité des signaux audio et vidéo restitués sur la *qualité d'expérience* du spectateur.

CHAPITRE III – ACTIVITE PHYSIOLOGIQUE ET OCULAIRE

Les mesures physiologiques et oculaires sont utilisées dans divers domaines de recherche comme des indicateurs objectifs par exemple des émotions humaines comme la colère, la tristesse ou la joie (Ekman, Levenson et Friesen, 1983). Le domaine de la recherche en facteurs humains utilise plutôt ces mesures pour déterminer l'effort mental ou la fatigue (Vicente, Thornton et Moray, 1987). Ces mesures présentent différents avantages par rapport aux mesures subjectives. Premièrement, elles ne sont pas assujetties aux différents biais de l'évaluation subjective (évaluation explicite, échelles, comparaison inter-séquences, *etc.*). Un de leur principal atout est de pouvoir mesurer de manière non invasive l'influence de la qualité, c'est-à-dire que la qualité n'est pas consciemment évaluée contrairement aux mesures subjectives. Par ailleurs, ces mesures n'impliquent pas de processus mnésiques et n'interfèrent donc pas avec le processus de traitement de l'information. En ce sens, les indicateurs physiologiques et oculaires sont qualifiés de mesures *objectives*.

Comme indiqué dans le chapitre I, la durée des séquences de test recommandées par l'UIT est fixée à quelques secondes afin d'éviter l'influence des biais mnésiques de récence et de primauté. L'enregistrement *on-line* (en continu) des mesures physiologiques et oculaires offre l'avantage de pouvoir observer l'influence des fluctuations de qualité au moment où elles se passent et ce, avec un niveau de précision temporelle élevée. Cette instantanéité, qui n'est pas capturée par les mesures subjectives (notes de qualité principalement recueillies après la phase de visualisation), permet d'envisager des contenus plus longs où la qualité fluctue comme cela pourrait être le cas en contexte réel de diffusion audiovisuelle. L'intégration de ces mesures permettrait ainsi de proposer un contexte d'évaluation plus représentatif d'un usage domestique.

Les indicateurs permettant d'accéder à l'activité physiologique puis ceux relatifs à l'activité oculaire seront détaillés dans ce chapitre. Leurs mesures et traitements statistiques seront également présentés. Cependant, avant d'aborder ces aspects, une description générale du fonctionnement du système nerveux humain sera apportée pour une meilleure compréhension des indices utilisés.

3.1. SYSTEME NERVEUX HUMAIN : REGISSEUR DE L'ACTIVITE PHYSIOLOGIQUE

Le système nerveux humain se divise en deux grands sous-systèmes complémentaires : le système nerveux central (SNC : encéphale et moelle épinière, centre du traitement des informations) et le système nerveux périphérique (SNPq : ensemble de fibres nerveuses parcourant le corps et relayant les informations sensorielles ou les commandes de fonction entre le SNC et les organes ou les muscles). Ce dernier se décline en deux nouveaux sous-systèmes, le système nerveux somatique (SNsoma) et le système nerveux autonome (SNA).

Le SNsoma transmet au SNC l'information provenant des récepteurs sensoriels (peau, ouïe, vue, *etc.*) et des propriocepteurs situés dans les articulations et les muscles. Il achemine

également les influx depuis le SNC jusqu'aux muscles striés squelettiques qui assurent la mobilisation (muscles oculaires par exemple), la stabilité et le maintien du corps. Les réponses motrices produites peuvent être déclenchées consciemment, ainsi, l'activité du SNsoma est dite volontaire.

Le SNA est également appelé système nerveux végétatif ou involontaire en raison de sa fonction de régisseur de l'ensemble des processus végétatifs (c.-à-d. le fonctionnement organique vital) comme par exemple, la régulation du rythme cardiaque, de la température corporelle ou de la motilité viscérale. Le SNA n'est donc pas contrôlé de manière volontaire d'où son appellation privilégiée de système nerveux autonome. Ce dernier innerve les glandes, les muscles lisses et cardiaque et le tractus gastro-intestinal. Le rôle du SNA est de maintenir l'homéostasie (maintien de la stabilité de l'environnement interne) et d'adapter l'activité physiologique pour fournir une réponse adaptée aux demandes environnementales (Critchley, 2002). Il est notamment impliqué dans un grand nombre d'activités réflexes et dépend d'influx sensitifs (relai de l'information sensitive viscérale ou interne au SNC) et d'efflux moteurs (relai des commandes de fonction du SNC aux organes ou aux muscles de l'activité végétative comme le muscle cardiaque). Plus précisément, le SNA est à l'origine de l'allocation ou de l'économie (hypoactivité) d'énergie pour pouvoir réagir de manière adaptée aux modifications de l'environnement.

3.1.1. PRINCIPE D'ACTIVATION PHYSIOLOGIQUE

L'organisme s'adapte donc aux modifications de l'environnement grâce à un système d'allocation/économie d'énergie régi par le SNA. Le terme « énergie » est généralement utilisé pour englober tous les mécanismes de dépenses énergétiques qui régulent l'organisme et influencent directement ou indirectement les processus physiologiques et psychologiques. Ce terme est choisi parce qu'il n'a pas de connotations théoriques spécifiques et peut éviter les confusions associées à des notions couramment utilisées telles que l'arousal, l'effort, la fatigue ou l'activation qui ont des significations spécifiques (Hockey, Coles et Gaillard, 1986, cité par Backs et Boucsein, 2000). Cependant, la plupart des chercheurs décrivent leurs résultats psychophysiologiques sur la base des fluctuations du niveau d'activation physiologique général (Backs et Boucsein, 2000).

En 1915, Cannon postulait que des états émotionnels comme la colère ou la peur étaient à l'origine d'une augmentation des dépenses énergétiques de l'organisme, cette conception a depuis été étendue à l'ensemble des conduites. Selon Pribram et McGuinness (1975), l'organisme aurait besoin d'atteindre un niveau d'activation physiologique suffisant pour réaliser une action. Chaque comportement serait impossible en-dessous d'un niveau spécifique d'activation mais pourrait être perturbé au-dessus (Lindsley, 1951). Par exemple, les performances motrices ou cognitives obtenues pour une tâche sont corrélées avec le niveau d'activation. Cependant, cette interaction ne présenterait pas une évolution monotone mais plutôt une courbe en U renversé (Hebb, 1955). Yerkes et Dodson (1908) ont en effet montré que les performances seraient maximales (maximum de ressources de traitement) pour un

niveau moyen d'activation physiologique. Lorsque le niveau d'activation dépasse cet optimum (stress), les performances diminuent. Cette loi s'appliquerait surtout aux tâches perceptivo-motrices simples et répétitives et serait moins généralisable aux tâches complexes (Collet, Roure, Rada, Dittmar et Vernet-Maury, 1996). Les variations du niveau d'activation dépendraient aussi de facteurs tels que les émotions, la fatigue ou encore l'effort.

De manière générale, le concept d'activation tente d'expliquer la relation entre les variations du niveau d'activité physiologique et les modifications du comportement (entendu au sens large du terme c.-à-d. incluant des activités telles que l'attention, la perception, *etc*.). La description du comportement à un moment donné requiert de considérer l'objectif vers lequel il est dirigé et son intensité (Duffy, 1972 ; Hebb, 1955). C'est cette notion d'intensité qui est communément appelée « activation » (parfois appelé *arousal*) ou « éveil » physiologique et pouvant être reflétée par le niveau de réponse d'un certain nombre de variables physiologiques. Par exemple, l'augmentation de la fréquence cardiaque ou de la sudation cutanée est liée à une activation accrue tandis que leur diminution indiquerait une baisse de l'activation. L'activation correspond à tous les processus dont l'objectif est de maintenir l'organisme prêt à agir (Pribram et McGuinness, 1975).

Les concepts d'éveil et d'activation peuvent cependant être distingués (Barry Clarke, McCarthy, Selikowitz et Rushby, 2005 ; Pribram et McGuinness, 1975 ; VaezMousavi, Barry, Rushby et Clarke, 2007), ces deux notions présentant des caractéristiques différentes des aspects énergétiques de la réponse physiologique et comportementale.

Pour Pribram et McGuinness (1975), le concept d'éveil renverrait à des phénomènes phasiques (réponse physiologique à court terme en réaction à un stimulus spécifique : évènements soudains ou modifiant l'environnement, Stern R., Ray et Quigley, 2001) tandis que l'activation serait associée à des phénomènes toniques (état précédant l'influence d'une stimulation, Stern R. *et al*., 2001). Stern R. *et al.* illustrent la notion de niveau tonique en prenant l'exemple de deux patients présents dans une salle d'attente, l'un doit subir une importante opération dentaire, l'autre est présent pour une consultation de routine. Le niveau tonique de l'activité physiologique de chacun des deux patients pourrait, par exemple, révéler un rythme cardiaque très élevé (130 battements par minute -bpm-) pour le premier et « normal » (80 bpm) pour le second. Pribram et McGuinness distinguent également un troisième mécanisme énergétique dit d'*effort* qui coordonnerait les deux premiers.

Boucsein (1993) sur la base des travaux de Pribram et McGuiness (1975) a différencié trois types d'activation physiologique (voir fig. 3.1). Le premier renvoie à l'activation générale de l'organisme et à « l'effort » et peut être rapproché de l'activation de Pribram et McGuinness. L'état d'activation peut être observé par les modifications toniques (non spécifiques) des mesures de l'activité du SNA. Le second correspond à l'activation dite affective (focalisation de l'attention : sur un nouvel évènement par exemple) pouvant être rapprochée de la notion d'éveil de Pribram et McGuiness et traduite par des modifications phasiques de l'activité physiologique. Le dernier type d'activation est dit de préparation à l'action ou dirigé vers un

but, ses corrélats physiologiques correspondent à une augmentation de l'activité physiologique (comme une augmentation de la fréquence cardiaque par exemple) anticipatrice de l'action. Les deux premiers types d'activation, présentés ci-avant, participeraient au système d'effort de Pribram et McGuiness. Les différents types d'activation proposés par Boucsein sont représentés dans la Figure 3.1 ci-dessous.

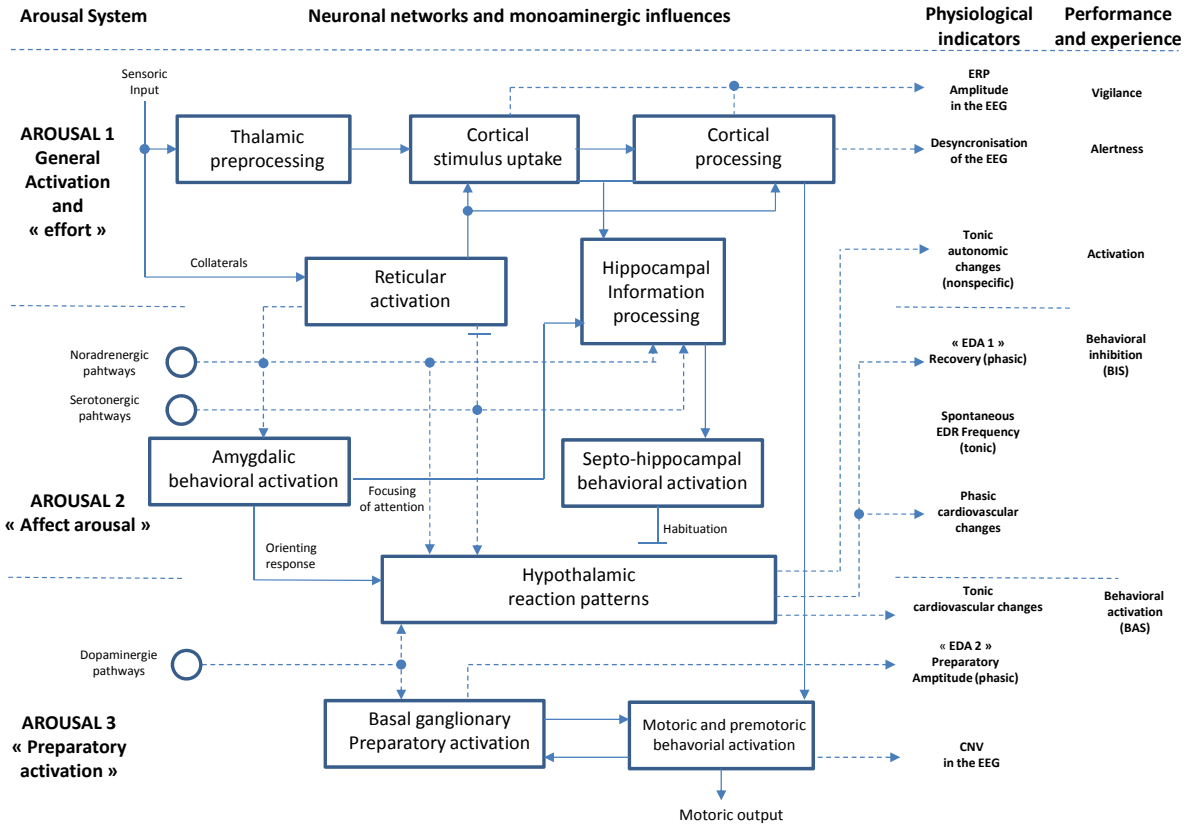


Fig. 3.1. Illustration des trois types d'activation physiologique du modèle de Boucsein (1993).

Barry *et al.* (2005) désignent l'éveil comme l'état énergétique du moment (non spécifique) tandis que l'activation se rapporterait à la mobilisation énergétique relative à la réalisation d'une tâche. Selon ces auteurs, les performances à la tâche seraient corrélées au niveau d'activation (réponse tonique). Pour VaezMousavi *et al.* (2007), une personne serait « activée » lors de l'exécution d'une tâche par rapport à une situation de non-tâche (appelée *baseline*). Cette activation peut alors être mesurée par la différence entre les niveaux toniques enregistrés pendant la baseline (condition de repos sans tâche ou tâche simplifiée, voir sect. 3.5 ci-dessous) et durant la tâche. Comme Barry *et al.*, ces auteurs considèrent l'éveil comme l'état énergétique du moment tandis que l'augmentation de l'état énergétique de base est définie comme l'activation observée à travers les variations physiologiques toniques recueillies entre la baseline et la tâche en cours.

Comme le précise Clarion (2009), la distinction entre éveil et activation n'est pas toujours évidente en raison de la proximité de ces concepts. Néanmoins, la notion d'activation peut se rapporter à une dimension tonique intensive (augmentation des dépenses énergétiques),

utilisée pour qualifier le niveau d'activité de l'organisme mesuré entre différentes conditions (repos *vs.* activité par exemple).

L'activation est le résultat d'une augmentation de l'activité du système nerveux central associée à une augmentation des processus nerveux périphériques (Eysenck, 1976, Lindsley, 1951). Elle est alors observable à travers, notamment, des indices du système nerveux autonome (Eason et Dudley, 1970). En effet, l'adaptation de l'organisme grâce au système d'allocation/économie d'énergie est régulée par l'activité de deux sous-systèmes antagonistes du SNA. L'un est impliqué dans les processus d'allocation (activation physiologique) tandis que le second est plutôt tourné vers l'économie d'énergie (repos, relaxation, digestion, *etc.*).

3.1.2. PARTITION DU SNA

Le SNA se divise en deux sous-systèmes : le système nerveux sympathique (SNS) et le système nerveux parasympathique (SNP). Ces deux systèmes dits antagonistes innervent l'ensemble de l'organisme et leur équilibre se fait par un processus d'activation/inhibition : lorsque l'un des deux systèmes est activé, l'autre est inhibé. Cependant, ce mécanisme n'est jamais absolu (pas de suppression totale de l'activité d'un des deux systèmes), ainsi, il est plus juste de parler de domination d'un système sur l'autre en réaction à des modifications exogènes (environnementales) ou endogènes (mécanismes mnésiques par exemple).

Le SNS est plutôt impliqué dans la dépense d'énergie et la préparation à l'action. L'activation du SNS (par libération de noradrénaline) se traduit par une dilation de la pupille (mydriase), une augmentation du rythme cardiaque, du débit sanguin, de la sudation cutanée ou encore par la déviation de l'afflux sanguin vers les organes vitaux et les muscles squelettiques. L'activation du SNS a également pour effet le ralentissement de la fonction intestinale et rénale.

Lors d'une situation stressante, le débit sanguin va augmenter pour favoriser une redistribution de l'irrigation sanguine (transport du sang facilité par la vasodilatation - dilatation des vaisseaux sanguins -) en faveur des organes de défense comme le cœur (avec un accroissement de la fréquence et de la force des battements), les muscles squelettiques ou encore le cerveau. Cette redistribution sanguine permet un apport sanguin accru, favorisant ainsi l'apport d'oxygène et de glucose, en direction des organes vitaux et des muscles actifs. L'apport d'oxygène et de glucose va permettre de préparer le corps à l'action dite de « combat » ou de « fuite ». Cette déviation sanguine en faveur des organes de défense signifie que le débit sanguin est réduit aux extrémités (vasoconstriction périphérique), comme le doigt par exemple. Le schéma (fig. 3.2) ci-dessous offre une illustration globale des différentes branches du système nerveux humain.

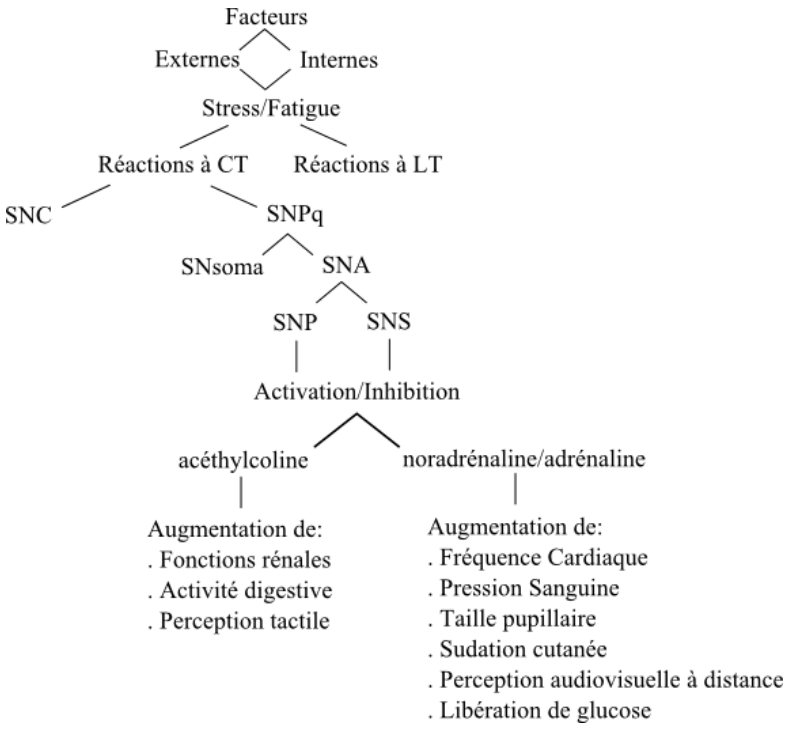


Fig. 3.2. Schéma des principales divisions du système nerveux humain et des conséquences physiologiques de l'activation des sous-systèmes nerveux parasympathique (SNP) et sympathique (SNS). Des stimulations à court terme (CT) ou à long terme (LT) vont entraîner des réponses du système nerveux central (SNC) et périphérique (SNPq). Les informations issues des récepteurs sensitifs comme la peau, les articulations, *etc.* sont relayées au SNC par le système nerveux somatique (SNsoma) dont les fibres motrices activent les muscles squelettiques sous contrôle volontaire du SNC. Le système nerveux autonome (SNA) relaie des informations de l'environnement interne au SNC ainsi que des informations du SNC aux organes et muscles impliqués dans l'activité végétative de l'organisme.

Le SNP quant à lui est plutôt stimulé dans les situations de repos et d'hypoactivité. Les conséquences physiologiques d'une activation du SNP (par libération d'acétylcholine) présentent le pattern inverse aux effets du SNS présentés ci-dessus. Par exemple, le SNP a tendance à freiner le rythme cardiaque et à augmenter la fonction rénale, l'activité digestive ou encore les perceptions tactiles. Le Tableau 3.1 ci-après récapitule les conséquences physiologiques de l'activation des systèmes nerveux sympathique et parasympathique.

Tableau 3.1. Récapitulatif des effets de l'activation du système nerveux sympathique (SNS) et parasympathique (SNP) où FC désigne la Fréquence Cardiaque (adapté d'après Widmaier, Raff et Strang, 2012).

Organes effecteurs	Action du SNS	Action du SNP
Yeux :		
Iris[1]	Dilatation (mydriase)	Rétrécissement (myosis)
Muscle ciliaire[2]	Relâche (aplatit le cristallin pour vision de loin)	Contracte (cristallin plus arrondi pour vision de près)
Cœur	Augmentation de la FC	Diminution de la FC
Artérioles cutanées	Constricte	Pas d'action
Veines	Constricte	Pas d'action
Estomac, Intestin	Diminue	Augmente
Peau :		
Glandes sudoripares[3]	Sudation	Pas d'action
Muscle de l'érection pilaire	Contracte	Pas d'action

Le SNA est en permanence sollicité et peut être modulé par différentes influences telles que des états émotionnels comme la peur, la joie ou la colère (Cacioppo, Berntson, Larsen, Poehlmann et Ito, 2000), des processus attentionnels (Boucsein, 2012 ; Lacey, J. et Lacey, B., 1970 ; Lang A., 1990 ; Porges, 1995) ou encore d'effort mental (Collet, Petit, Champely et Dittmar, 2003, cité par Clarion ; Fishel, Muth et Hoover, 2007 ; Grossman, Stemmler et Meinhart, 1990 ; Kahneman, 1973 ; Kahneman, Tursky, Shapiro et Crider, 1969 ; Porges, 1992 ; Sanders, 1990). Ces aspects seront plus amplement détaillés dans le chapitre IV suivant.

Ainsi, une domination du SNS serait plutôt rapprochée d'un état d'activation (effort, stress) tandis qu'une domination du SNP serait plutôt attribuée à un état de repos (lors d'un état de fatigue par exemple). Il existe aujourd'hui de nombreux outils permettant de mesurer les modifications de l'activité du SNA.

1 Contrôle de la taille pupillaire

2 L'action du muscle ciliaire contrôle la courbure du cristallin (lentille suspendue au corps ciliaire, derrière l'iris) permettant l'accommodation pour la vue de près ou de loin. Le muscle ciliaire se contracte pour l'accommodation de près (forme arrondie du cristallin) et se relâche pour la vision de loin (aplatissement du cristallin). Cet effort de déformation permet de maintenir une image nette de l'environnement indépendamment d'une certaine distance.

3 Voir section 3.3 ci-après

3.2. INDICES DE L'ACTIVITE CARDIAQUE

3.2.1. MECANISME DU FONCTIONNEMENT CARDIAQUE

La fréquence cardiaque est un indicateur reconnu du niveau d'activation générale. Par exemple, une augmentation du rythme cardiaque survient au cours d'une activité physique, d'une activité sexuelle ou d'un effort mental (Frijda, 1986). Une coupe détaillée du cœur est apportée par la Figure 3.3.

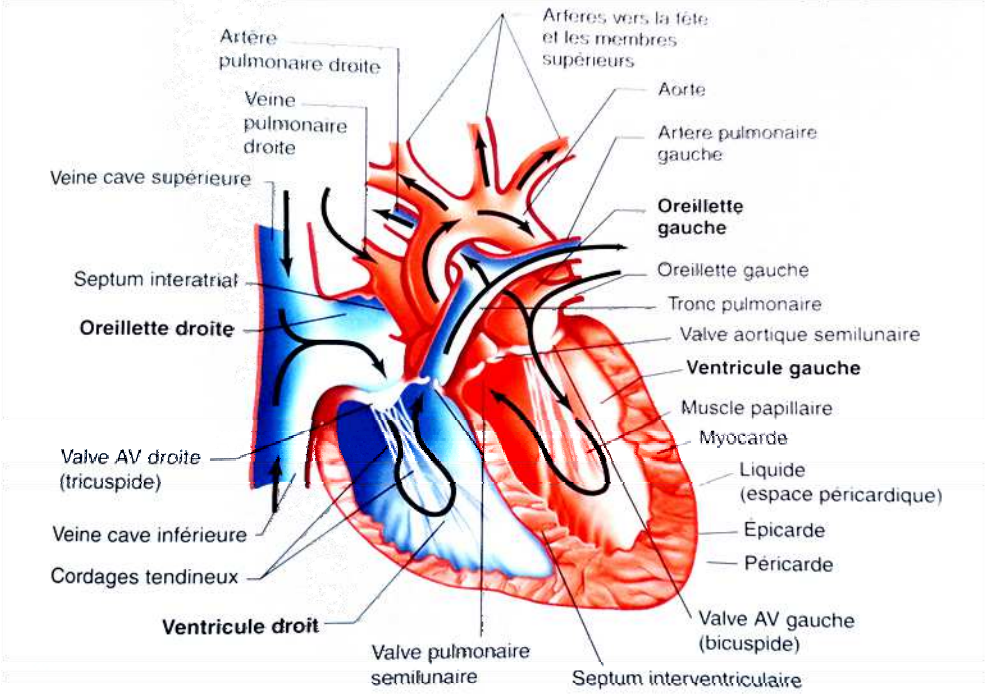


Fig. 3.3. Coupe détaillée du cœur, les flèches indiquent l'écoulement sanguin (d'après Widmaier *et al.*, 2012).

Le cœur est un organe musculaire dont la taille atteint environ celle d'un poing fermé. Du point de vue anatomique, il se divise en deux parties, gauche et droite, chacune composée de deux compartiments : l'oreillette et le ventricule. La circulation du sang de l'oreillette au ventricule se fait grâce à la valve dite atrio-ventriculaires. Un cycle cardiaque (un battement) est défini par la contraction et la décontraction de ces compartiments grâce à la contraction initiale du myocarde, paroi du cœur constituée de cellules musculaires. Les contractions du cœur agissent comme une pompe double : les parties gauche et droite du cœur pompent séparément et simultanément le sang. A chaque battement, le cœur pousse le sang *via* l'aorte vers le réseau artériel pour acheminer oxygène et nutriments vers les différents organes et muscles. Le retour sanguin (appauvri en oxygène et nutriments) vers le cœur s'effectue *via* le réseau veineux (grande circulation). Afin de redevenir du sang artériel « propre », le sang veineux est propulsé par le ventricule droit, *via* l'artère pulmonaire, vers les poumons pour revenir au cœur dépouillé de son gaz carbonique et riche en oxygène (petite circulation). Ainsi traité, il pourra à nouveau être éjecté pour irriguer les différents tissus organiques.

Une contraction cardiaque correspond à une série complexe d'évènements électriques impliquant une succession de polarisation et dépolarisation des cellules musculaires atriales et ventriculaires. Le potentiel d'action à l'origine de la contraction est initié par le nœud sino-atrial qui agit comme un stimulateur (pacemaker) naturel permettant de contrôler les cycles de contractions cardiaques. La fréquence de sa décharge correspond à la fréquence cardiaque soit le nombre de contractions cardiaques par minute. L'influx se propage ensuite vers le nœud atrio-ventriculaire pour permettre l'éjection sanguine.

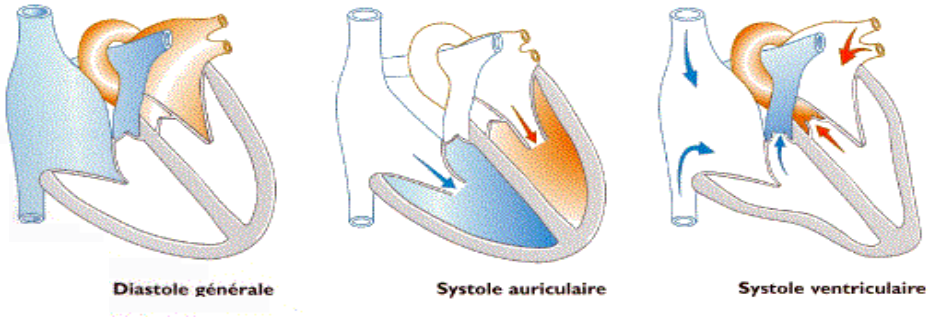


Fig. 3.4. Révolution cardiaque (d'après Lacombe, 2009). Les zones colorées représentent le parcours sanguin dans chaque ventricule, la couleur bleu représente le ventricule droit et le rouge le ventricule gauche.

Plus précisément, les phases de contraction auriculaire et ventriculaire sont appelées systoles, la phase de relaxation correspond à la diastole. La phase systolique se divise en deux temps (voir fig. 3.4 ci-dessus). Tout d'abord, la systole auriculaire correspondant à la contraction des oreillettes (1/10s). Cette contraction évacue le sang contenu dans les oreillettes vers les ventricules qui vont à leur tour se contracter (3/10s). On parle alors de systole ventriculaire. Durant la systole ventriculaire, le myocarde se contracte et la valve atrio-ventriculaire se ferme. Dans un second temps, les valves aortiques et pulmonaires s'ouvrent et l'éjection ventriculaire du sang commence. Après cette étape, se produit la diastole qui correspond au repos cardiaque (4/10s) et pendant laquelle les valves atrio-ventriculaires s'ouvrent pour permettre au sang d'affluer de nouveau vers les oreillettes grâce au retour veineux. En parallèle, les valves aortiques et pulmonaires se ferment. Une révolution cardiaque complète aura donc une durée d'environ 8/10 de seconde. Le cœur sain se contracte environ 60 à 120 fois par minute (variabilité selon le niveau sportif, l'âge, l'état émotionnel, *etc.*). Ces contractions vont permettre au cœur de pomper en moyenne 70 ml de sang par contraction soit environ cinq litres de sang par minute. Cette quantité peut augmenter jusqu'à 25 l en cas d'efforts intenses. Cette adaptation est possible grâce au principe de domination sympathique ou parasympathique.

3.2.2. INFLUENCE DU SYSTEME NERVEUX AUTONOME

Le cœur, comme de nombreux organes, est doublement innervé à la fois par le système nerveux parasympathique et sympathique. Comme précédemment indiqué, ces deux sous-systèmes du SNA produisent des effets opposés sur l'activité cardiaque.

3.2.2.1. INFLUENCE PARASYMPATHIQUE

Le SNP exerce en permanence une action de ralentissement du tonus cardiaque *via* le nerf vague (principale innervation parasympathique du cœur), on parle alors de système cardio-modérateur. Cette décélération cardiaque serait associée également aux comportements d'approches, aux phénomènes attentionnels ainsi qu'aux *inputs* informationnels (Porges, 1995).

3.2.2.2. INFLUENCE SYMPATHIQUE

Le SNS à une action accélératrice du rythme cardiaque traduite notamment par une augmentation de la fréquence cardiaque. On parle alors de système cardio-accélérateur. Cette accélération cardiaque serait associée à une activation physiologique pour une préparation générale de l'organisme à l'action ainsi qu'à la mobilisation de certains types de ressources (Obrist, 1981).

L'activité du rythme cardiaque est donc régulée, en fonction des besoins de l'organisme, par l'adaptation des rôles accélérateur et ralentisseur des systèmes nerveux sympathique et parasympathique.

3.2.3. MESURE

La mesure du rythme cardiaque est classiquement réalisée au moyen d'un électrocardiogramme (ECG). L'ECG permet l'enregistrement des modifications de l'activité électrique pendant le cycle cardiaque. Cependant, la fréquence cardiaque (FC) telle qu'étudiée dans ce document a été obtenue à partir d'un enregistrement par *pléthysmographie*. Cette technique d'enregistrement consiste à détecter les variations de volume sanguin périphérique (VSP) grâce à un petit boitier placé à l'extrémité du doigt. La réflexion d'un rayon infrarouge permet d'enregistrer la quantité de lumière renvoyée. En effet, le sang va absorber le spectre lumineux et ne renvoyer que la couleur rouge. Donc, plus le volume sanguin est important, plus il y aura de lumière rouge réfléchie. Cette technique d'enregistrement permet d'étudier les variations du volume sanguin aux extrémités corporelles (doigts, lobe de l'oreille, orteil). Face à une situation impliquant une domination de l'activité sympathique, le VSP va diminuer en raison de la déviation sanguine vers les organes de défenses (cœur, muscles, organes vitaux). En d'autres termes, le VSP diminue face à une situation impliquant une activation du SNS.

3.2.4. CAPTEUR ET SITE D'ACCUEIL

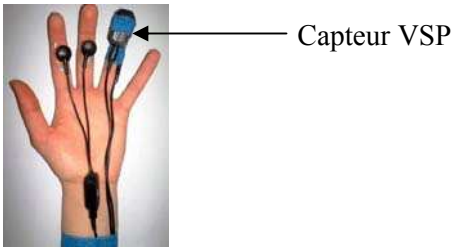


Fig. 3.5. Site d'accueil du pléthysmographe pour la mesure du VSP et de la FC.

Le capteur VSP se présente comme un petit boitier (environ 2,5 cm) que l'on place sur le côté palmaire du doigt (fig. 3.5). Il est généralement maintenu par une sangle ou un adhésif. Il est toutefois important de ne pas trop serrer l'adhésif afin de ne pas gêner la circulation sanguine ce qui pourrait atténuer ou aplatir l'onde de pouls.

Le capteur VSP est un capteur très sensible (mouvement, passage de la lumière, problème de fixation) notamment au mouvement qui est une des principales sources d'artefacts du signal. Les mouvements peuvent être à l'origine de battements manqués ou de battements supplémentaires. Ainsi, il est impératif de demander au participant de bouger le moins possible la main ou le bras. Afin de ne pas induire de gêne (agacement, fatigue) imputable à cette immobilité, il convient de proposer un environnement confortable et de déterminer par avance la position la plus agréable pour le participant afin que celle-ci puisse être maintenue durant l'expérimentation.

Enfin, il est nécessaire de prêter attention à la température périphérique du participant à son arrivée. En effet, une température trop basse (doigts froids), pouvant être l'expression de la nervosité du participant ou l'impact résiduel d'une faible température extérieure, traduit une circulation réduite aux extrémités, c'est-à-dire une mesure biaisée au départ. Cela est vrai autant pour l'enregistrement du VSP que pour celui de la réponse électrodermale et de la température cutanée périphérique qui seront abordées dans la suite de ce chapitre. Deux solutions complémentaires permettent de minimiser ce phénomène si non pathologique : premièrement, il faut prévoir un temps d'adaptation pour le participant (réchauffement général, diminution de la nervosité) et deuxièmement, il convient de maintenir une température ambiante favorable à l'enregistrement des mesures physiologiques (aux alentours de 23°).

3.2.5. SIGNAL

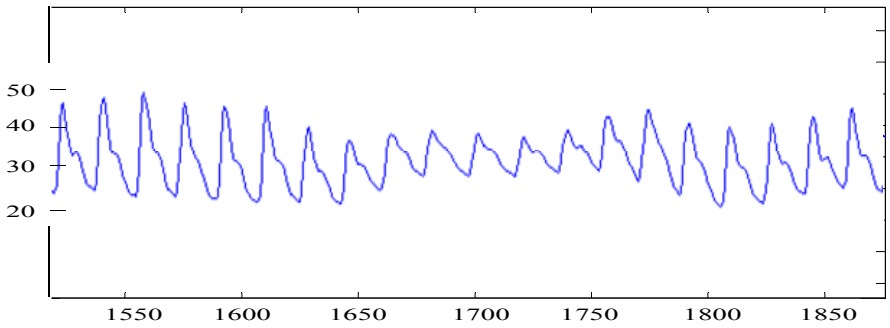


Fig. 3.6. Extrait d'un tracé de mesure du VSP.

La quantité de lumière réfléchie est transformée en signal électrique. Le signal obtenu correspond à une mesure relative qui n'a pas d'unité standard. La fréquence cardiaque va ensuite pouvoir être calculée à partir des mesures du VSP.

La montée correspond à la contraction systolique, la pente suivante sera plus lente que la montée observée. Par ailleurs, dans certains cas, le signal peut mettre en évidence un rebond lors de la descente (encoche dicrote) en raison de la fermeture de la valve aortique. La pression descend ensuite régulièrement jusqu'à la systole suivante. Un exemple de tracé du VSP est donné par la Figure 3.6 ci-avant. L'intervalle crête à crête du signal augmente ou diminue en fonction de l'activation sympathique.

3.2.6. TRAITEMENT DES MESURES

Les différents indicateurs statistiques permettant l'étude de l'activité cardiaque sont appelés indicateurs de la variabilité du rythme cardiaque (VRC). La VRC au cours du temps peut être divisée en deux classes (Lantelme, Custaud, Vincent et Milon, 2002) : la variabilité à court terme (quelques secondes à quelques minutes, détection des changements rapides de la FC) et la variabilité à long terme (plusieurs heures). L'étude de la VRC, notamment à court terme, consiste en l'analyse des variations des intervalles entre deux battements cardiaques (intervalles RR) ou de la fréquence cardiaque. Les indices statistiques de la VRC peuvent être calculés soit dans le domaine temporel soit dans le domaine fréquentiel. La liste, non exhaustive, des indicateurs utilisés dans ces deux domaines pour l'étude de la variabilité à court terme est présentée ci-après.

3.2.6.1. DOMAINE TEMPOREL

Différents indicateurs peuvent être utilisés pour étudier la VRC dans le domaine temporel. Tout d'abord, la moyenne de la fréquence cardiaque constitue un indicateur valide et usité (Boonnithi et Phongsuphap, 2011; Souza Neto, Neidecker et Lehot, 2003). Il est à noter que cet indicateur de tendance centrale est sensible aux valeurs extrêmes à l'origine d'une plus grande variabilité. D'autres indicateurs descriptifs comme l'étude des valeurs minimums et maximums peuvent également apporter de l'information pertinente. L'étude de l'écart-type

peut aussi s'avérer utile pour étudier la dispersion de l'échantillon étudié. D'autres indices peuvent être calculés dans le domaine temporel à partir des intervalles RR ou à partir de la comparaison des différences entre les intervalles RR adjacents. La *Task Force* (1996) recommande notamment l'étude des trois indicateurs suivants :

- **SDNN :** déviation standard de l'intervalle RR (ms) sur toute la période d'enregistrement, renseigne sur la variabilité à court terme,
- **SDANN** : déviation standard de la moyenne des intervalles RR (ms) de segments de cinq minutes sur toute la période d'enregistrement, renseigne sur la variabilité globale des segments (variabilité à long-terme),
- **RMSSD** : obtenu à partir de la différence entre intervalles RR (ms) et correspondant à la racine carrée des différences au carré des intervalles RR successifs, renseigne sur la variabilité d'origine parasympathique et modulée par la respiration (Souza Neto *et al.*, 2003).

Le calcul du SDNN est recommandé par la *Task Force* autant pour l'étude de la variabilité à long terme qu'à court terme. Le calcul du RMSSD est notamment recommandé pour l'étude de variabilité à court terme.

De manière conventionnelle, la VRC est étudiée à partir de segment de cinq minutes ou vingt-quatre heures. Cependant, plusieurs auteurs ont étudié la VRC à partir de segments beaucoup plus courts, validant ainsi cette approche (Boonnithi et Phongsuphap, 2011 ; Salahuddin, Cho, Jeong et Kim, 2007 ; Thong, Li, McNames, Aboy et Goldstein, 2003). Par exemple, Boonnithi et Phongsuphap (2011) suggèrent que les indicateurs de tendance centrale « moyenne de la fréquence cardiaque » et « moyenne des intervalles RR » sont les deux indicateurs pertinents du domaine temporel pour estimer l'influence du stress mental (induit par la réalisation d'un tâche de type *Stroop*) sur la VRC. Ce constat est issu de l'étude de segment temporel de cinquante secondes. En d'autres termes, l'étude de la moyenne, à partir de segments très courts, permet de discriminer des changements de la variabilité cardiaque. Cette observation permet d'affirmer que la mesure de l'activité cardiaque est valide pour rendre compte de l'activité physiologique du spectateur en présence de dégradations appliquées durant quelques secondes.

3.2.6.2. DOMAINE FREQUENTIEL

L'objectif principal de l'analyse de la VRC par une approche fréquentielle est d'obtenir des informations sur le rapport ente l'activité sympathique et parasympathique. Ce rapport est appelé index ou balance sympatho-vagale. Plus précisément, la balance sympatho-vagale permet d'obtenir un retour sur l'état d'activation dominante, pour un contexte et un individu donné, du SNS ou du SNP. Cette prédominance d'un système sur l'autre s'opère à la fois sur le niveau tonique et phasique.

Le spectre de puissance du rythme cardiaque est généralement décomposé en trois bandes de fréquences : **TBF** (Très Basse Fréquence entre 0,003 et 0,04 Hz), **BF** (Basse Fréquence

entre 0,04 et 0,15 Hz) et **HF** (Haute Fréquence entre 0,15 et 0,4 Hz). Une quatrième bande de fréquence, UBF (Ultra Basse Fréquence, <0,003 Hz), est également utilisée pour l'analyse d'enregistrements réalisés sur de longue période (24 h). Les TBF traduiraient des mécanismes de régulation à long terme liés à la thermorégulation, à la vasomotricité ou d'autres facteurs (Souza Neto *et al.*, 2003).

Plusieurs études ont montré que la distribution de la puissance et de la fréquence centrale des BF et HF ne sont pas fixes mais variables selon les changements des modulations du SNA (Malliani, Pagani, Lombardi et Cerutti, 1991 ; Pagani *et al.*, 1986).

La bande HF correspond à l'arythmie sinusale respiratoire (ASR) provoquée par l'activité mécanique des poumons. Ce qui est plus intéressant, selon Malliani (1999), est que la bande HF serait également un bon indicateur de l'activité parasympathique de l'organisme.

A l'inverse, bien que ce résultat soit sujet à controverse, la bande de fréquence BF serait un marqueur de l'activité sympathique (Kamath et Fallen, 1993 ; Malliani *et al.*, 1991; Montano, Ruscone, Porta, Lombardi, Pagani et Malliani, 1994 ; Pagani *et al.*, 1986 ; Pagani, Furlan, Pizzinelli, Crivellaro, Cerutti et Malliani, 1989). La bande BF comporterait tout de même une composante traduisant une activité parasympathique (Akselrod, Gordon, Ubel, Shannon, Berger et Cohen, 1981 ; Appel, Berger, Saul, Smith et Cohen, 1989). La représentation de l'activité sympathique par la bande BF serait améliorée par son expression en unité normalisée (Malliani *et al*., 1991) soit :

(Puissance de la bande de fréquence étudiée / par la totalité du spectre) × 100

Une autre méthode pour refléter la balance sympatho-vagale est l'utilisation des BF et HF normalisées (*n*) suivant une des formules suivantes (Malliani, Lombardi, Pagani et Cerutti, 1990) :

BFn= 100 × BF/ (HF + BF ou (+TBF))
ou HFn= 100 × HF/ (HF + BF ou (+TBF))
ou BF/HF

Le rapport BF/HF serait notamment pertinent pour l'étude des modulations sympathiques (Malliani, 1999).

Devant la quantité des indicateurs proposés, Boonnithi et Phongsuphap (2011) préconisent l'étude des BF normalisées (BFn, augmente lors d'une domination sympathique), de la différence entre les BF et HF normalisées (dBFHF, augmentation lors d'une dominance sympathique) et du rapport BF/HF appelé SVI (Sympathovagal Balance Index, augmentation lors d'une dominance sympathique) dans le cadre de l'étude de segments d'observation de courtes durées. Ces trois indicateurs peuvent être rassemblés sous l'appellation commune de ratios. Le Tableau 3.2 ci-après présente, pour chaque ratio, le détail de son calcul.

Tableau 3.2. Calcul des ratios BFn, dBFHF et SVI utilisés pour l'étude de la variabilité du rythme cardiaque (selon Boonnithi et Phongsuphap, 2011).

Ratios	Unités	Calcul
BFn	%	100 x BF / (HF + BF+ TBF)
dBFHF	%	| BFn – HFn |
SVI	%	BF/HF

D'autres méthodes comme les méthodes géométriques (Task Force, 1996) ou le calcul de la réactivité cardiaque (approche phasique) ont été utilisées pour décrire la VRC. Cependant, ces méthodes ne seront pas abordées dans ce document.

3.3. INDICES DE L'ACTIVITE ELECTRODERMALE

L'activité électrodermale (AED) est connue pour être un indice fiable des variations d'activation physiologique (Collet *et al.*, 1996) et notamment pour refléter l'activation du SNS (Boucsein, 2012 ; Dawson, Schell et Filion, 2007). L'AED est un des indicateurs les plus utilisés dans le domaine de la psychophysiologie (définie dans le chap. IV). En effet, elle est reconnue pour réagir fortement aux stimuli psychologiques signifiants (Boucsein, 2012 ; Stern R. *et al.*, 2001) tels que les fluctuations émotionnelles (Féré, 1988), les processus de prise de décision (Bechara, Damasio, H., Damasio, A. et Lee, 1999), les états de stress ou encore l'activité cognitive (Siddle, 1991). La mesure de l'AED s'effectue généralement au moyen d'une mesure exosomatique consistant en l'application d'un léger voltage entre deux électrodes pour recueillir le niveau de conductance cutanée.

3.3.1. FONCTIONNEMENT DE L'ACTIVITE ELECTRODERMALE

La peau, véritable interface entre l'organisme et son environnement, joue un rôle de senseur du monde extérieur grâce aux nombreuses terminaisons sensorielles dont elle est le siège. Elle a également un rôle de protection contre la lumière, les infections et sa composition perméable protège l'organisme des fluides extérieurs (Clarion, 2009). Par ailleurs, elle participe également à la thermorégulation de l'organisme grâce aux phénomènes de sudation et de vasoconstriction ou vasodilatation. L'apparition de sueur à la surface de la peau s'effectue grâce à l'activité de glandes appelées glandes sudoripares. La densité de ces glandes, environ trois millions réparties inégalement sur le corps humain, est maximale sur les sites plantaires, palmaires et sur le front. Il existe deux types de glandes sudoripares : apocrines et eccrines. L'AED est le résultat de l'activité des glandes sudoripares eccrines et survient soit en réaction à des besoins de régulations thermiques (effort, chaleur), soit de façon réflexe à un niveau plus local (piqûre, électricité, chaleur) mais elle peut également refléter certains états psychologiques comme des états émotionnels ou de stress (Boucsein, 2012, p.31). Cette sudation « psychologique » est principalement observée sur la plante des pieds et la paume des mains. L'activité des glandes sudoripares eccrines peut être mesurée grâce à l'enregistrement des variations des propriétés électriques de la peau.

3.3.2. INFLUENCE DU SYSTEME NERVEUX AUTONOME

Il est communément admis que les glandes sudoripares eccrines sont uniquement innervées par le système nerveux sympathique (Dawson *et al.*, 2007). L'AED présente donc la particularité d'être sous le contrôle exclusif du SNS, ce qui facilite son interprétation. En cas de domination sympathique, les glandes sudoripares eccrines sont activées engendrant ainsi une micro-sudation cutanée. La production de sueur par les glandes eccrines sera d'autant plus élevée que l'activation sympathique sera importante. L'AED reflète donc les variations de sudation et peut être mesurée grâce au niveau de conductance ou de résistance cutanée lorsqu'un léger voltage est appliqué. Plus précisément, la conductance de la peau va augmenter en cas d'activation sympathique. Une illustration familière de l'activité électrodermale en cas d'activation sympathique est l'expérience des mains moites en cas de stress important ou de nervosité. Diverses explications ont été proposées pour expliquer l'augmentation de sudation lors d'une activation du SNS. Un premier postulat considère que cette action durcit la peau, la protégeant contre d'éventuelles lésions (Wilcott, 1967). Il a en effet été observé que la peau est plus difficile à couper durant une forte sudation (Edelberg et Wright, 1962). Une deuxième théorie suppose que la sudation permettrait de refroidir l'organisme pour favoriser la préparation à une éventuelle activité de combat ou de fuite. Pour Darrow (1937), la fonction de la sudation cutanée serait principalement de fournir une surface adhésive pour améliorer l'acuité tactile et l'adhérence aux objets.

3.3.3. MESURE

La mesure de l'AED consiste à enregistrer les variations des propriétés électriques de la peau, c'est-à-dire l'activité des glandes sudoripares eccrines. Cette mesure s'effectue grâce à l'application d'un courant électrique continu de faible intensité (imperceptible pour le participant) passant par deux électrodes. Ce dispositif permet d'établir un circuit électrique où le participant devient la résistance variable. Ce processus permet de mesurer la variation de la conductance cutanée, en temps réel, engendrée par les changements de sudation. Une autre mesure consiste à mesurer la résistance cutanée. Dans ce document, les variations d'AED sont étudiées au moyen des mesures de la conductance cutanée.

3.3.4. CAPTEUR ET SITE D'ACCUEIL

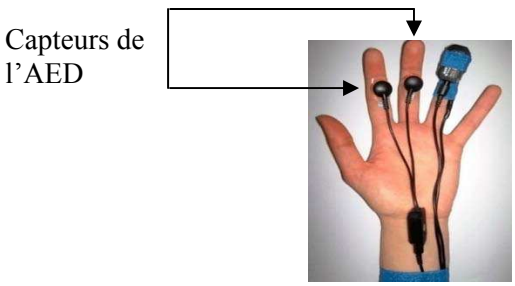


Fig. 3.7. Site d'accueil pour les capteurs de mesure d'AED correspondant à deux électrodes installées ici sur les phalanges adjacentes médiales de l'index et du majeur.

Afin de mesurer le niveau de conductance cutanée, deux petites électrodes sont apposées sur des sites présentant une densité importante de glandes sudoripares eccrines, généralement les faces palmaires ou plantaires.

La mesure de la conductance cutanée est dépendante de la surface de contact entre l'électrode et le site d'accueil (taille de la zone de contact, lieu) en matière de niveaux mesurés et d'amplitudes. A titre indicatif, une aire de contact de 1.0 cm^2 est recommandée lorsque le site le permet (Fowles, Christie, Edelberg, Grings, Lykken et Venables, 1981). Afin d'optimiser cette mesure, Fowles *et al.*, (1981) recommandent l'usage de disques adhésifs double-face pour permettre à la fois un meilleur maintien de l'électrode et un meilleur contact avec la peau. Par ailleurs, le placement des électrodes est bi-site et doit être effectué sur les phalanges distales ou médiales des doigts adjacents d'une même main (Fowles *et al.*, 1981 ; Venables et Christie, 1980). La phalange médiale présente notamment l'avantage d'être moins sensible aux mouvements, une des principales sources d'artefacts des mesures de l'AED. En conséquence, les mouvements de la main sur laquelle les capteurs sont installés doivent être limités. La Figure 3.7 donne un exemple de l'installation des électrodes pour la mesure d'AED.

Ces auteurs recommandent également l'utilisation d'électrodes en argent et chlorure d'argent (Ag-AgCl) pour l'enregistrement de la conductance cutanée. Ce type d'électrode doit être utilisé avec un gel électrolytique isotonique facilitant la continuité électrique entre la peau et l'électrode.

Pour ne pas altérer la qualité du signal, il convient de ne pas nettoyer la peau avec de l'alcool avant la pose des électrodes. En effet, cette solution a pour effet de diminuer la résistance électrique cutanée (et donc d'augmenter la conductance, Boucsein, 2012).

L'AED est également influencée par la température extérieure (Venables et Christie, 1980), cette dernière devra, dans la mesure du possible être maintenue aux alentours de 23°C (Boucsein, 2012). Par ailleurs, pour les mêmes raisons que celles mentionnées pour la mesure du rythme cardiaque, il est important de laisser un temps d'adaptation au participant avant le début de l'enregistrement. Il est préférable que les électrodes soient installées dès l'arrivée du participant afin de le familiariser avec le port des capteurs.

3.3.5. SIGNAL

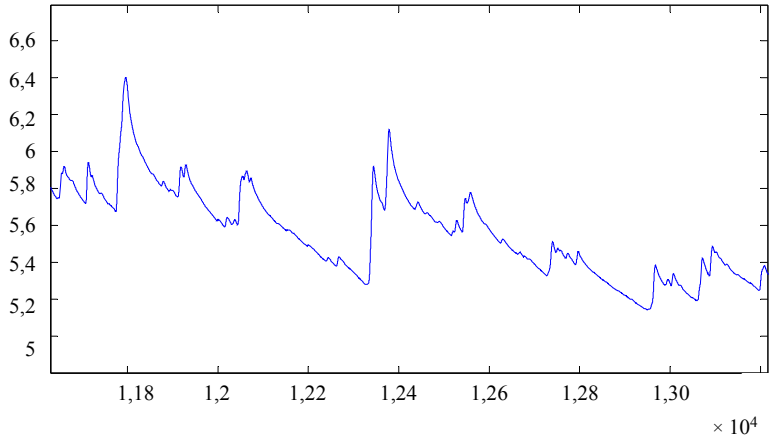


Fig. 3.8. Extrait d'un tracé de mesure d'AED.

L'unité de mesure standard de la conductance est le Siemens (S). La conductance de la peau est mesurée en micro-Siemens (µS). La conductance électrodermale est généralement comprise entre environ 2 µS, lorsque l'individu est détendu, et 20 µS. Cependant, ce niveau varie considérablement en fonction de facteurs environnementaux et/ou inter (type de peau) ou intra-individuelles (Dawson *et al.*, 2007). Le type de signal enregistré par le capteur de conductance cutanée présente généralement une croissance rapide et une baisse relativement lente. Cependant, de manière générale, le niveau de conductance électrodermale tend à diminuer au cours du temps en situation de repos. Un exemple de tracé d'AED est donné par la Figure 3.8.

3.3.6. TRAITEMENT DES MESURES

L'AED présente deux phénomènes distincts, un phénomène tonique (niveau de l'activité électrodermale au fil du temps, NED) et un phénomène phasique (réaction ou réponse électrodermale -RED- localisée dans le temps, en réaction à un évènement ou un stimulus) (Boucsein, 2012). Lors de la présentation d'un stimulus, des RED peuvent être observées au sein de l'AED. Ces réactions transitoires, de durées limitées, sont également observées en l'absence d'évènement particulier, on parle alors de réponse électrodermale non spécifique (RED-NS). L'ensemble de ces réponses NED, RED et RED-NS est illustré par la Figure 3.9 ci-dessous.

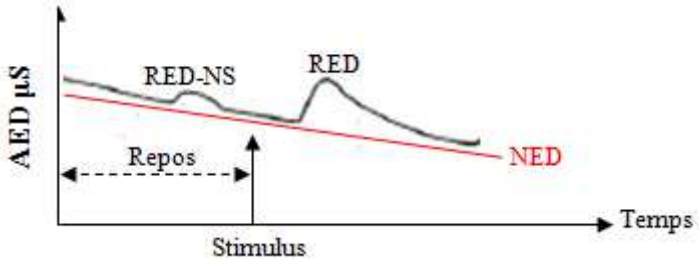


Fig. 3.9. Illustration des états toniques (NED) et phasiques (RED et RED-NS) de l'AED (adapté d'après Clochard, 2011).

Les traitements statistiques relatifs à l'approche tonique sont abordés dans le paragraphe suivant, ceux relatifs à l'approche phasique peuvent être consultés dans l'annexe 3-A.

APPROCHE TONIQUE

Le niveau électrodermal (NED) permet de quantifier les variations lentes de l'AED et peut être mesuré grâce au niveau de conductance électrodermale. Le niveau tonique est sensible aux modulations sympathiques et est fréquemment utilisé pour l'évaluation de l'activation physiologique (Barry et Sokolov, 1993) : une augmentation serait corrélée avec une augmentation de l'activation (Hastrup, 1979 ; Sostek, 1978). Les variations toniques de l'AED sont reconnues pour être l'indice autonome le plus fiable pour mesurer les variations d'activation (Woodworth et Schlosberg, 1954). Celles-ci peuvent être observées entre une condition de référence (repos, activité particulière) et une condition expérimentale. La méthode la plus répandue pour l'étude du niveau tonique consiste à calculer la valeur moyenne du NED pour une fenêtre temporelle donnée. Cette méthode suppose de connaitre par avance les segments temporels à étudier, l'objectif étant de pouvoir comparer entre elles les différentes fenêtres. La taille de la fenêtre ne sera pas sans impact sur la moyenne obtenue. Selon Clarion (2009), une période suffisamment longue permettra de diminuer l'effet des RED mais diminuera également la comparabilité entre segments. Boucsein (2012) préconise des segments d'une durée de 10 à 30 secondes. En dehors de la moyenne, Clarion (2009) propose également l'étude de la dispersion, augmentant lors de l'apparition de RED, pour refléter l'activité phasique.

3.4. INDICE DE TEMPERATURE CUTANEE PERIPHERIQUE

Les artérioles cutanées sont innervées par le système sympathique et parasympathique. Cependant, comme indiqué dans le Tableau 3.1 (sect. 3.1.2 ci-avant), l'activation parasympathique a un effet négligeable tandis que l'activation sympathique est à l'origine d'une vasoconstriction des vaisseaux cutanés (rétrécissement).

3.4.1. INFLUENCE DU SYSTEME NERVEUX AUTONOME

La température cutanée périphérique (TCP) mesurée aux extrémités du corps (doigt, orteil) constitue un autre indicateur de l'activité du SNA puisqu'elle varie en fonction de l'irrigation sanguine de la peau. La TCP dépend, entre autre, de l'activation du système sympathique. Plus précisément, la TCP diminue lors d'une domination sympathique. Cette diminution s'explique par un phénomène de vasoconstriction périphérique conséquent à une réduction du débit sanguin aux extrémités en raison de la déviation sanguine vers les organes de défense. La mesure de la TCP permet d'obtenir les réponses des vasoconstricteurs modulées par le SNS (Kistler, Mariauzouls et von Berlepsch, 1998).

3.4.2. MESURE

L'enregistrement de la TCP est réalisé grâce à un capteur de petite taille (thermistor) fixé au doigt. Les changements de température enregistrés sont ensuite convertis en variation de courant électrique.

3.4.3. CAPTEUR ET SITE D'ACCUEIL

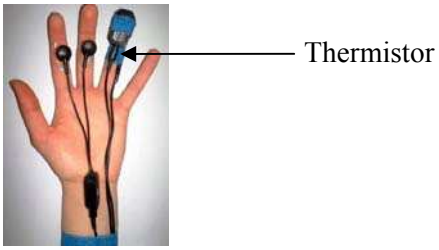


Fig. 3.10. Site d'accueil du capteur de mesure des variations de TCP.

Le capteur de TCP peut être attaché sur la face dorsale ou palmaire de n'importe quel doigt (fig. 3.10 ci-dessus) ou orteil. Le capteur peut être maintenu à l'aide d'une sangle ou d'un adhésif.

3.4.4. SIGNAL

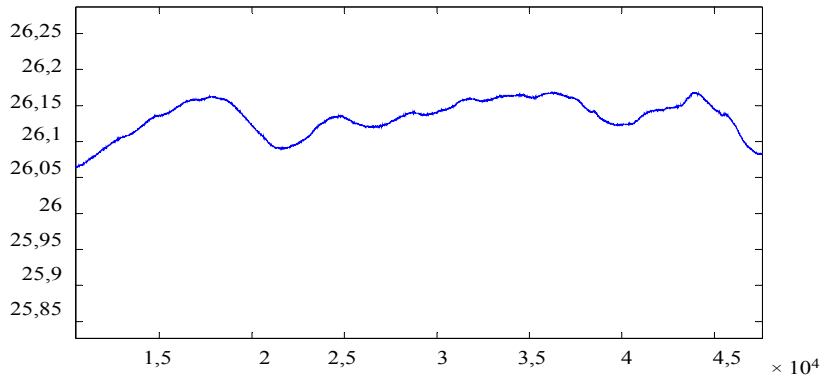


Fig. 3.11. Extrait d'un tracé de mesure de TCP.

L'unité de mesure utilisée pour les expérimentations décrites dans ce document est le degré Celsius (°C). Les variations de TCP sont assez faibles, de l'ordre de quelques centièmes de degré (1 à 10 centièmes dans les cas les plus extrêmes, Averty, 1998, cité par Clarion, 2009). La TCP est donc un indice aux variations lentes et de faibles amplitudes. Ainsi, l'échelle du signal recueilli présentait une granularité fine (échelle de valeurs relativement petite) afin de favoriser l'étude des variations de température au cours du temps. La température périphérique, telle que mesurée au doigt, se situe généralement aux alentours de 28°C. Un exemple de signal recueilli sur plusieurs minutes est apporté par la Figure 3.11 ci-dessus.

Les spécifications des capteurs (liées au matériel utilisé pour l'enregistrement des mesures) décrits ci-dessus sont présentées dans le Tableau 3.3.

Tableau 3.3. Spécification des indices de Température Cutanée Périphérique (TCP), de Volume Sanguin Périphérique (VSP) et de l'Activité ElectroDermale (AED *via* conductance cutanée).

Indices	Type de capteurs	Unité de mesures	Spécifications
TCP	Thermistor SA9310M	Degré Celsius (C°)	Limites : 10°C-45°C Précision : ± 1°C
VSP	HR/BVP Flex/Pro SA9308M	Unité moins quantité affichée de 0% à 100%	Précision : ± 5%
AED	Flex/Pro SA9309M ou V91-01, 8mm, Coulbourn	microSiemens (µS)	Limites : 0 à 30 µS Précision : ± 0,2 µS

3.5. DIMINUTION DE LA VARIABILITE INTERINDIVIDUELLE

De manière générale, les mesures physiologiques sont soumises à une grande variabilité interindividuelle. Afin de pouvoir comparer entre elles les mesures issues de différents individus, notamment les valeurs de NED mais également l'ensemble des mesures physiologiques enregistrées (FC, VSP et TCP), il convient de rendre les valeurs étudiées comparables. Pour cela, le signal peut être normalisé selon une valeur de référence enregistrée, la plupart du temps, avant la présentation des conditions expérimentales. Cette période est appelée *baseline* et est généralement mesurée à l'état de repos. La moyenne des mesures enregistrées durant la baseline sera ensuite utilisée comme valeur de référence individuelle pour les signaux enregistrés pour chaque participant. Les moyennes toniques de chaque signal enregistré seront ensuite divisées par cette référence individuelle. Cette étape de normalisation des données permet d'effectuer des comparaisons interindividuelles. La durée de la baseline peut être variable, cependant, selon Gerin, Pieper et Pickering (1994) un enregistrement de cinq minutes serait suffisant pour obtenir une mesure stable. L'enregistrement de la baseline à l'état de repos sous-tend que le participant n'est impliqué dans aucune tâche, son activité n'est donc pas contrôlée. Certains auteurs rapportent un état de somnolence ou d'anxiété lié à l'attente de la tâche à venir (Farha et Sher, 1989, cité par Jennings, Kamarck, Stewart, Eddy et Johnson, 1992). Il est également envisageable que les participants puissent s'ennuyer ou penser à des évènements plus ou moins positifs durant la phase de repos. Cela pourrait biaiser la *baseline* en tant que condition de comparaison. Fishel *et al.* (2007) rajoutent que l'utilisation d'une baseline mesurée au repos (c.-à-d. mesurant un faible niveau d'activation physiologique) augmenterait la détection de niveaux d'activation élevée et diminuerait la sensibilité de l'analyse à détecter les niveaux d'activation faible.

Pour pallier ces problèmes, Jennings *et al.* (1992) ont suggéré la mesure d'une *vanilla* baseline définie comme la réalisation d'une tâche cognitive, similaire à la tâche expérimentale, mais requérant un niveau d'effort cognitif moins important (par exemple, si la tâche expérimentale nécessite la réalisation de calculs mentaux complexes, la *vanilla* baseline pourrait demander la réponse à des calculs simples). Cette *vanilla* baseline permettrait de

diminuer les biais induits par une mesure enregistrée lors d'un faible niveau d'activation physiologique. Elle présenterait également l'avantage de maintenir la vigilance du participant durant l'enregistrement et d'activer les processus qui seront engagés dans la tâche principale.

3.6. INDICATEURS OCULAIRES

D'un point de vue général, l'œil se présente comme un globe rempli de liquide et protégé de trois parois distinctes divisibles en deux chambres : la chambre antérieure (entre l'iris et la cornée, remplie d'humeur aqueuse pour le maintien de la pression intra-oculaire) et la chambre postérieure (entre le cristallin et la rétine, remplie d'humeur vitrée, soit 90% du volume de l'œil, pour le maintien de la rigidité de l'œil). Une description plus complète de l'anatomie de l'œil est présentée dans l'annexe 2-A. La mobilité du globe oculaire est assurée par la coordination de six muscles externes présentés dans la Figure 3.12 ci-dessous.

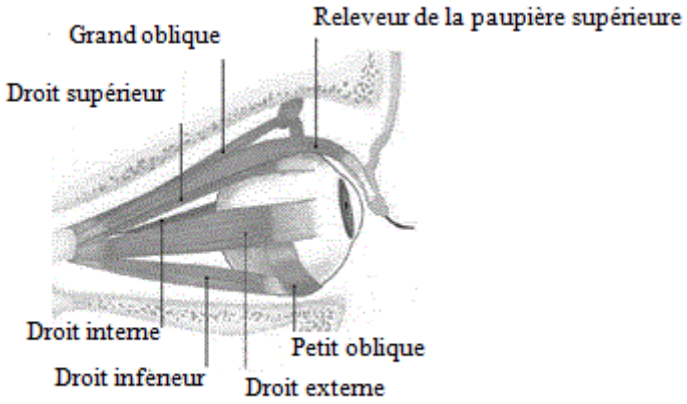


Fig. 3.12. Les différents muscles moteurs du globe oculaire (d'après le Lacombe, 2009).

Ce corpus musculaire permet de réaliser des mouvements selon trois axes : transversal (haut/bas), antéro-postérieur (intorsion/extorsion) et vertical (temporal/nasal). Les différents mouvements oculaires peuvent être regroupés selon leurs caractéristiques physiques (lents ou rapides), réflexes (réflexes ou volontaires) ou directionnels (conjonctifs ou disjonctifs). Le déclenchement de ces mouvements permet de maintenir une cible fixe (saccade ou vergence) ou mobile (poursuite oculaire) sur la fovéa. Les mouvements oculaires sont aussi parfois des réactions réflexes aux mouvements de l'observateur ou de l'environnement (réflexe optocinétique ou vestibulo-oculaire, pour plus de détails sur les différents mouvements oculaires se référer à l'annexe 2-A). Le clignement de paupière est réalisé grâce à l'action du muscle releveur de la paupière supérieure dont la fonction physiologique est de protéger l'œil contre les agressions extérieures.

L'étude du comportement oculaire correspond principalement à l'observation des différents mouvements des yeux. Ce sont ces mouvements qui vont permettre une perception exacte du monde extérieur. En effet, rapides et précis, ils maintiennent un point de fixation (zone pertinente de l'image) sur la rétine et plus particulièrement sur la fovéa (zone située au centre de la rétine) où l'acuité est maximale. Par exemple, l'activité de lecture implique un

déplacement constant de l'œil pour maintenir en zone fovéale les lettres devant être lues (empan visuel d'environ trois lettres de chaque côté du point de fixation). Ainsi pour permettre la vision (fovéale ou périphérique), les yeux ne sont jamais immobiles. Ces mouvements oculaires permettent au système visuel d'acquérir des informations en analysant les aspects pertinents de l'environnement.

A l'exception des muscles lisses *ciliaires* (permettant l'accommodation du cristallin) et de *l'iris* (contrôle du diamètre pupillaire) doublement innervés par le SNP et le SNS, les muscles oculomoteurs sont des muscles striés squelettiques innervés par les nerfs crânien III, IV et VI (pour les mouvements oculaires, Andreassi, 2007 ; Stern R. *et al.*, 2001 ; Widmaier *et al.*, 2012) et VII (pour le clignement : Kramer, 1991 ; Morris et Miller, 1996) dépendant de l'activité du système nerveux somatique (Widmaier *et al*., 2012). L'étude de l'activité oculaire regroupe donc différents types d'indicateurs : des indicateurs du comportement (mouvements et clignements) et des indicateurs du système nerveux autonome (diamètre pupillaire, accommodation).

La mesure de l'activité oculaire est réalisée au moyen de techniques d'oculométrie. Celles-ci permettent d'obtenir un retour de l'activité oculaire d'un individu à travers différents indices comme les fixations, les saccades, les clignements des yeux ou encore le diamètre pupillaire. Ces indicateurs sont aujourd'hui reconnus pour être des indices valides d'effort mental ou de fatigue induits par une tâche (voir Andreassi, 2007 ; Kramer, 1991 ; Stern R. *et al.*, 2001). Typiquement, les indices relatifs aux clignements ou à la fermeture de l'œil sont fréquemment utilisés pour évaluer les phénomènes de fatigue ou de baisse de vigilance dans des contextes aussi différents que la conduite automobile, le pilotage d'avion, le contrôle aérien, l'utilisation et la gestion de machines industrielles, *etc*.

Les différents indicateurs oculaires utilisés dans les expérimentations présentées dans ce document sont détaillés dans les paragraphes suivants. Il est à noter que les descriptions données (durées, paramètres, calculs, *etc.*) sont celles relatives au matériel utilisé à savoir l'outil de mesure oculaire faceLAB™ (SeeingMachine). Certains paramètres ne font pas l'objet de consensus (taille de fenêtre de calcul ou durée de fixations par exemple) et sont donc fonction des spécifications intrinsèques à l'outil de mesure. Ainsi, pour plus de simplicité, le principe de fonctionnement de l'outil de mesure, commun à l'ensemble des indices recueillis, est apporté avant la présentation des indicateurs oculaires.

3.6.1. MESURE DU COMPORTEMENT OCULAIRE

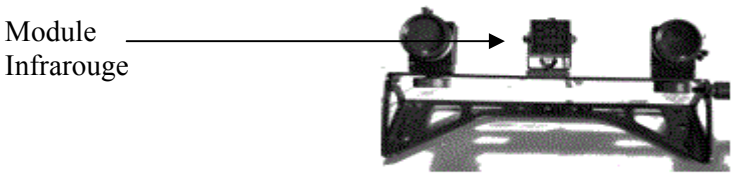


Fig. 3.13. Présentation du hardware faceLAB pour l'enregistrement de mesures oculaires.

faceLAB™ est un outil de mesure oculométrique (eye tracking) développé par SeeingMachine. L'outil (eye tracker), constitué de deux petites caméras mobiles et d'un module infrarouge (voir fig. 3.13 ci-dessus), est installé face au participant à une distance dépendante de la taille de l'écran. Les caméras vont permettre d'enregistrer en temps réel (enregistrement d'un flux vidéo et traitement parallèle des images capturées) des indices du comportement oculaire d'un individu. La technique d'enregistrement consiste à détecter la position des yeux grâce au reflet cornéen d'une lumière infrarouge (diodes) dirigée vers le centre de la pupille. Le rayon émis va en effet être renvoyé grâce aux propriétés réflectrices de la cornée puis enregistré par les caméras. Le rayon lumineux permet également d'amplifier la brillance de la pupille pour faciliter sa détection par la caméra. Cette technique est appelée vidéo-oculographie. Le couplage de la brillance de la pupille, du reflet cornéen et d'un système de traitement de l'image va permettre de détecter la position spatiale de l'œil et le calcul du diamètre pupillaire en quasi-permanence.

Les deux caméras sont placées entre le participant et l'objet d'interaction (un écran par exemple) et installées de manière à converger vers le visage du participant. Chacune des caméras a une fonction définie. La première zoome sur les yeux du participant tandis que la seconde filme le visage dans sa globalité. Le reflet cornéen détecté par les caméras est intégré dans un référentiel de coordonnées afin de permettre une reconstruction de la position et de l'orientation de la tête du participant dans l'environnement ainsi que la position et la direction du regard.

Pour une détection correcte du regard par l'outil de mesure, un modèle de tête individuel doit être créé pour chaque participant. La création de ce modèle consiste à détecter plusieurs points de référence sur le visage du participant. Ces points correspondent aux coins extérieurs et intérieurs de l'œil et aux commissures des lèvres. faceLAB va également retenir d'autres points d'appui pour lesquels les zones de contrastes ne devraient pas être modifiées sous l'influence de rotations de la tête ou de mouvements faciaux (coins du nez, sourcils, *etc.*) Le maillage établi va constituer un point d'ancrage qui permettra à l'*eye tracker* de repérer en permanence la pupille et ce, malgré la présence de mouvements de la tête, tout au moins dans une certaine mesure (une rotation de la tête jusqu'à 30 degré est tolérée). La création d'un modèle individuel contribue à établir un environnement d'évaluation plus écologique où le participant garde une certaine liberté de mouvement (pas de mentonnière). Le suivi (tracking) de pupille par faceLAB est présenté par la Figure 3.14 ci-dessous.

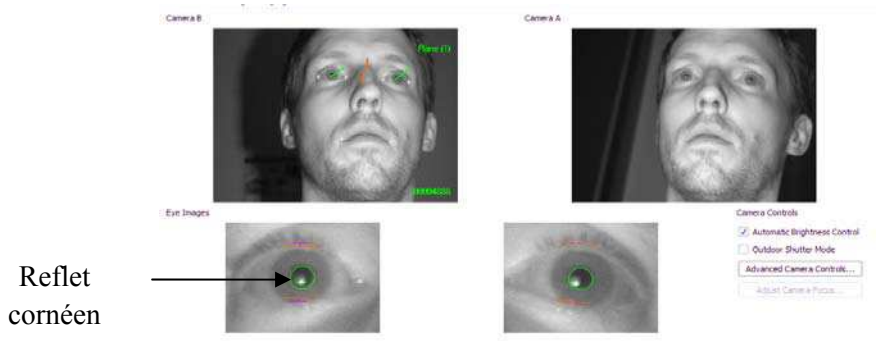


Fig. 3.14. Interface de faceLAB pour le contrôle du tracking de pupille.

Une phase de calibrage, pendant laquelle il est demandé au participant de suivre des yeux un point se déplaçant à différents endroits de l'écran (jusqu'à neuf emplacements pour un calibrage optimal), est ensuite réalisée. Cette phase permet d'optimiser, pour chaque participant, la précision avec laquelle le regard sera détecté autour de la zone de fixation (compensation entre l'endroit d'apparition du point et l'endroit où est mesuré le regard). Un bon calibrage est indispensable lorsque l'étude des cheminements oculaires durant la visualisation est souhaitée.

La configuration complète de l'*eye tracker*, de la création du modèle de tête (sélection des points de références par détection automatique et correction manuelle) à la phase de calibrage, nécessite en moyenne quinze à vingt minutes selon la facilité de l'outil à détecter les points d'ancrages, la qualité du calibrage des caméras, selon que l'individu soit porteur de lunettes ou non, *etc*.

faceLAB propose l'extraction d'un grand nombre d'indicateurs oculaires. Dans ce document, les indices du diamètre pupillaire, de saccades, de la fréquence et la durée des clignements de paupières et de pourcentage de fermeture de l'œil (PERCLOS) ont été étudiés.

3.6.2. DIAMETRE PUPILLAIRE

La pupille (ouverture permettant l'entrée de la lumière) varie sous l'action de deux muscles lisses antagonistes de l'iris contrôlés par le SNA : le sphincter dont la contraction permet le rétrécissement de la pupille (innervation parasympathique) et le dilatateur dont la contraction dilate la pupille (innervation sympathique). Le rétrécissement (myosis) de la pupille serait la conséquence d'une augmentation de luminosité (réflexe photo-moteur), d'une vue rapprochée (réflexe d'accommodation) et plus généralement d'une activation parasympathique (liée à un phénomène de fatigue par exemple). A l'inverse, la dilation de la pupille (mydriase) serait associée à une baisse de luminosité, une vue de loin ou une activation sympathique (liée à une activation physiologique) (Menche et Schäffler, 2004, p. 210). La taille de la pupille peut varier de 2 à 8 mm et se réduit naturellement avec l'âge. De petites variations du diamètre de la pupille (souvent inférieure à 0,5 mm) reflèteraient les processus cognitifs (Andreassi, 2007) tels que les processus attentionnels, émotionnels ou du traitement de l'information.

SIGNAL

La *pupillométrie* renvoie aux mesures des variations du diamètre de l'ouverture pupillaire de l'œil. Le signal correspond à la taille du diamètre pupillaire, en mètre, calculée par faceLAB à chaque image enregistrée (60 images par seconde).

Le diamètre pupillaire reflète donc l'activité du SNA, en revanche, les indices présentés ci-après correspondent à des indices du comportement oculaire relevant d'une innervation du système nerveux somatique.

3.6.3. SACCADES

Comme précédemment indiqué, une des principales fonctions des mouvements oculaires est de permettre aux yeux de modifier leur position pour se focaliser sur de nouvelles zones d'intérêt. Les mouvements de saccades permettent d'amener les informations pertinentes en vision fovéale (acuité visuelle maximale). Une *saccade* se définit comme un mouvement oculaire rapide (vélocité angulaire jusqu'à 800°/s selon Berthoz et Petit, 1996) permettant d'amener le regard d'une zone d'intérêt à une autre (une description plus détaillée est apportée dans l'annexe 2-A). Les saccades permettent notamment d'explorer une scène visuelle par l'alternance d'instants de fixation (permettant le traitement de l'information, Just et Carpenter, 1984) et de déplacements qui portent le regard jusqu'à un nouvel espace de fixation. Ces mouvements balistiques (pas de redirection possible une fois le mouvement enclenché) peuvent être d'amplitude variable selon que l'activité implique une fenêtre d'exploration large (balayage d'une pièce) ou réduite (lecture). Les saccades jouent un rôle important dans la construction des représentations visuelles de l'environnement (Rayner, 1998) et reflètent également les mouvements de l'attention visuelle (Remington, 1980). Ces mouvements inter-fixations sont également reconnus pour fournir des indices sur le niveau de fatigue (Bahill et Stark, 1975 ; Schmidt, Abel, DellOsso et Daroff, 1979 ; Stern J., Boyer, Schroeder, Touchstone et Stoliarov, 1994).

SIGNAL

facaLAB fournit un algorithme de détection des saccades s'appuyant sur un modèle physiologique. Les composantes de durée et de distance ne seront pas investiguées dans ce document. Le signal mesuré correspondait à l'information de présence de saccades permettant d'obtenir un retour sur un éventuel état de fatigue induit par les conditions expérimentales. Ainsi, la mesure extraite par faceLAB correspondait à une réponse binaire (présence ou absence de saccades) obtenue pour chaque image enregistrée. faceLAB détecte également avec précision les saccades même lors des clignements des yeux.

3.6.4. CLIGNEMENT DE PAUPIERES (BLINK)

Le *clignement de paupière* ou *Eye Blink* (EB) est un phénomène facilement observable du comportement et se définit comme la fermeture rapide des yeux suivie, dans un très court intervalle de temps, par une ré-ouverture. Un clignement correspond à un mouvement spontané et indispensable à l'intégrité de la cornée par son action d'humidification du film lacrymal. Cette fréquence subit une diminution avec l'âge (Mourant, Lakshmanan et Chantadisai, 1981). Le clignement est effectué grâce à l'action du muscle releveur de la paupière supérieure et dure environ 0,2 à 0,4 s. Il surviendrait environ 15 000 fois par jour (Tecce 1992) avec une moyenne de quinze à vingt fois par minute lorsqu'un individu est détendu. Tecce (1992) souligne qu'un adulte a besoin de seulement deux à quatre clignements par minute pour maintenir l'humidité nécessaire à l'intégrité de l'œil, la plupart des clignements ne sont alors pas nécessaires du point de vue physiologique.

Trois types de clignements sont généralement distingués (Orchard et Stern J., 1991) : réflexe (protection de l'œil contre une lumière vive ou de la poussière par exemple), volontaire (contrôle conscient) et endogène (associés au processus de traitement de l'information, attentionnels, d'effort mental, de fatigue, *etc.*).

Selon Ponder et Kennedy (1927, p.10, cité par Stern J., Boyer et Shroeder, 1994) « *la fréquence de clignement serait étroitement liée à la tension mentale d'un individu à un moment donné, ces mouvements constitueraient une sorte de mécanisme d'allégement de l'énergie nerveuse* ». La fréquence des EBs (EBfreq) refléterait des facteurs psychologiques comme l'état émotionnel ou l'effort mental. En effet, la mesure des caractéristiques des EBs comme la durée ou la fréquence est utilisée depuis plus de soixante ans pour étudier l'activité mentale (voir Kramer, 1991). Les EBs sont également reconnus pour être des indicateurs de fatigue (Kramer, 1991).

SIGNAL

faceLAB distingue les clignements rapides des yeux et les clignements plus longs associés à la détection de fatigue. Pour cet outil, un EB est défini par une durée n'excédant pas 0,35 s. Au-delà, le mouvement de paupière ne sera pas marqué comme EB. L'évènement de clignement peut être divisé en différentes composantes comme la durée ou la fréquence. faceLAB permet d'extraire ces deux composantes :

- **Fréquence (EBfreq) :** moyenne du nombre des EBs détectés pour une fenêtre temporelle d'une seconde,
- **Durée (EBdur)** : calcul de la durée moyenne des EBs détectés pour une fenêtre temporelle d'une seconde.

Les EBs réguliers (nécessité physiologique) sont également détectés. Ceux-ci sont notamment nécessaires au calcul d'autres indicateurs oculaires tels que le PERCLOS décrit ci-après.

3.6.5. PERCLOS

Le *PERCLOS* (percentage of eyelid closure over the pupil over time) est reconnu comme étant un indice fiable de fatigue (Dinges et Grace, 1998 ; Lang L. et Qi, 2008 ; Wierwille, Wreggit et Knipling, 1994). Il est fréquemment utilisé dans les études portant sur l'évaluation du niveau de vigilance comme pour la conduite sur route (fatigue du conducteur) par exemple.

SIGNAL

Le calcul du PERCLOS permet d'obtenir le pourcentage de fermeture de l'œil (couverture de l'iris par la paupière) pour un intervalle de temps donné. Généralement cet intervalle est compris entre une et cinq minutes (Sommer et Golz, 2010). La méthode de calcul ne tient pas compte des clignements courts et réguliers des paupières. Les valeurs de PERCLOS

correspondent au pourcentage d'images, pour une fenêtre temporelle donnée, pour lesquelles l'œil présente une fermeture d'au moins 75% (c.-à-d. que l'œil est recouvert à 75%). Le paramétrage par défaut, proposé par faceLAB, fixe la taille de la fenêtre d'intégration pour le calcul du PERCLOS à cent quatre-vingt secondes.

Une synthèse des spécifications des indices oculaires décrits est proposée par le Tableau 3.4 ci-après.

Tableau 3.4. Synthèse des spécifications des indices de clignement : durée (EBdur) et fréquence (EBfreq), de PERCLOS, de diamètre pupillaire (DP) et de saccades (SAC).

Indices oculaires	Unité de mesure	Description
EBdur	seconde	durée : moyenne par seconde
EBfreq	hertz	fréquence : moyenne par seconde
PERCLOS	valeur entre 0 et 1	% des images où l'œil est recouvert à 75% pour une fenêtre temporelle de 180s
SAC	booléenne	0 (absence) – 1 (présence)
DP	mètre	taille de la pupille (obtenue pour chaque image enregistrée)

3.7. VERS DES MESURES PSYCHOPHYSIOLOGIQUES

Les paragraphes précédents ont présenté le fonctionnement général du système nerveux périphérique (SNsoma et SNA) ainsi que les différents indicateurs permettant d'étudier son activité. Les indicateurs du système nerveux autonome correspondent à la fréquence cardiaque (FC), étudiée à travers le volume sanguin périphérique (VSP), la température cutanée (TCP), le diamètre pupillaire (DP) et l'activité électrodermale (AED). Cette dernière reflète uniquement l'activé du SNS tandis que les autres indices dépendent à la fois de l'activité du SNS et de celle du SNP. Les indices du comportement oculaire à savoir la durée et la fréquence des clignements des yeux (EBdur et EBfreq), la saccade (SAC) et le PERCLOS, sont principalement représentatifs de l'activité du système nerveux somatique (voir fig. 3.15 ci-après).

De manière générale, l'augmentation des indicateurs du SNS traduit un phénomène d'activation physiologique liée à une allocation énergétique accrue. L'allocation ou l'économie d'énergie (ou la domination du SNS ou du SNP) est soumise à de nombreuses influences. Par exemple, l'état émotionnel, les processus attentionnels ou les phénomènes d'effort mental ou de fatigue vont moduler le SNA. Différentes études ont montré que ces mêmes facteurs influencent aussi le comportement oculaire d'un individu.

Les indicateurs présentés ont en effet été choisis en raison de leur validité et de leur utilisation dans le cadre d'étude de la fatigue ou de l'effort mental (Andreassi, 2007 ; Kramer, 1991, Stern R. *et al.*, 2001). Les relations entre les processus mentaux et l'activité physiologique ou oculaire sont étudiées par le domaine de la psychophysiologie. Le chapitre suivant aborde ces aspects.

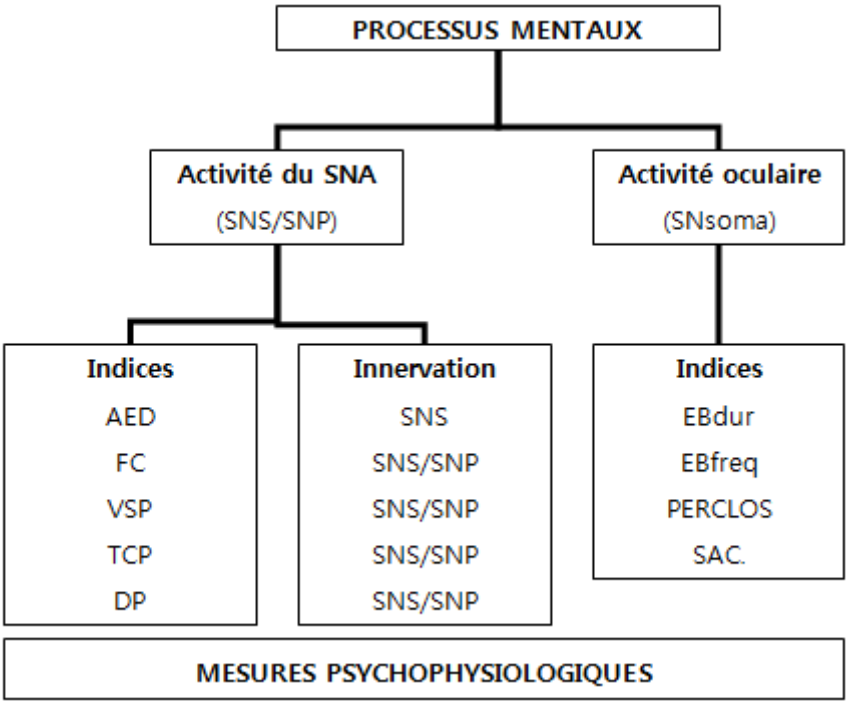


Fig. 3.15. Récapitulatif des indicateurs physiologiques et oculaires étudiés.

CHAPITRE IV – PSYCHOPHYSIOLOGIE

Les mesures physiologiques comme le rythme cardiaque ou la réponse électrodermale et celles de l'activité oculaire comme les clignements ou le degré de fermeture de l'œil sont employées depuis des années par la psychologie, puis plus récemment par l'ergonomie. Le domaine de la psychophysiologie s'intéresse à la mesure des réponses physiologiques et/ou oculaires lors d'activités aussi diverses que le sommeil, la résolution de problèmes, les réactions au stress, l'apprentissage, la mémoire, les processus de traitement de l'information, la perception, *etc.* Plus précisément, la psychophysiologie correspond à la branche de la psychologie qui étudie les changements de l'activité physiologique en réaction à des *inputs* psychologiques (Turner, 1994, cité par Ravaja, 2004). Selon Andreassi (2007), la psychophysiologie peut être définie comme « *l'étude des relations entre les manipulations psychologiques et les réponses physiologiques qui en résultent, mesurées dans l'organisme vivant, afin de promouvoir la compréhension des relations entre les processus mentaux et corporels* » (Andreassi, 2007, p. 28). Cacioppo et Tassinary (1990) précisent que : « *la psychophysiologie concerne l'étude des phénomènes cognitifs, émotionnels et comportementaux liés et révélés par les principes et évènements physiologiques* ». Les mesures psychophysiologiques permettent d'obtenir des index de l'activité cérébrale, à travers l'activité du SNA ou du comportement oculaire, selon les variations des processus attentionnels, émotionnels ou d'efforts moteur ou mentaux.

Dans le cadre de l'élaboration d'une méthode proposant d'étudier l'influence de la qualité sur le *coût utilisateur*, du point de vue de la fatigue ou de l'effort mental induit par un signal dégradé, il est nécessaire de connaitre la manière dont les indicateurs physiologiques et oculaires répondent à ces phénomènes. La visualisation d'un contenu suppose également l'implication d'autres processus capables d'influencer les mesures enregistrées. En effet, dans la vie de tous les jours, lorsqu'un individu visualise un contenu audiovisuel (film, documentaire, sport, *etc.*), les processus attentionnels et émotionnels mis en jeu influencent non seulement la perception générale du contenu mais aussi l'activité physiologique et oculaire du spectateur. Par exemple, des attributs du contenu audiovisuel comme le mouvement ou les changements de plans peuvent avoir un impact significatif sur la manière dont les spectateurs expérimentent et évaluent le contenu présenté.

Ainsi, différents facteurs autres que le niveau de qualité des signaux restitués sont à même d'influencer l'activité oculaire et physiologique d'un individu engagé dans une activité de visualisation. Dans le cadre de l'intégration de tels indicateurs aux méthodes d'évaluation de qualité, il convient de connaitre ces diverses influences.

4.1. MODULATIONS EMOTIONNELLES : VALENCE ET AROUSAL

L'expérience émotionnelle d'un individu est susceptible d'influencer le niveau de qualité perçue, comme une tolérance plus ou moins grande vis-à-vis des dégradations selon la nature de l'expérience émotionnelle et de moduler l'activité du SNA (James, 1884 ; Cannon, 1927)

et oculaire (Hess, 1972). Les réponses émotionnelles peuvent être subjectivement évaluées à travers les dimensions notamment de valence et d'arousal et potentiellement reflétées par les indicateurs psychophysiologiques.

Selon Wundt (1886) l'expérience subjective et physiologique des émotions peut être étudiée à partir de trois dimensions : la valence (plaisir *vs.* déplaisir), la vigilance ou dominance (tension *vs.* relâchement) et l'arousal (excitation *vs.* relaxation/calme). L'arousal est défini comme le niveau d'activation physiologique (c.à.d. l'intensité) associé à une réponse émotionnelle, d'un état de forte excitation à un état de calme ou de somnolence (Lang P., Greenwald, Bradley et Hamm, 1993). Lang P. (1980) a développé des échelles standardisées pour permettre l'évaluation de ces dimensions à partir de neuf catégories exprimées par des représentations picturales (échelles SAM, Self-Assessment Manikin). Une illustration de ces échelles est donnée par la Figure 4.1.

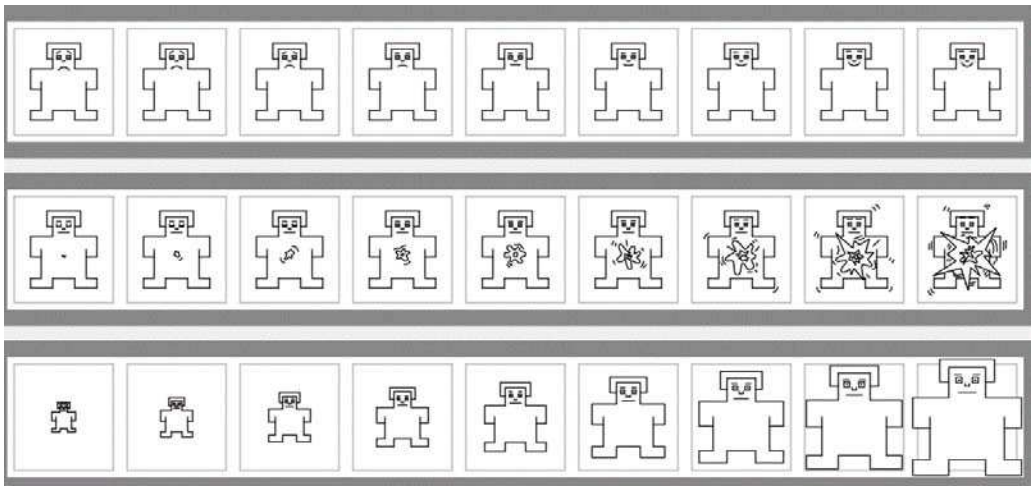


Fig. 4.1. Illustration des échelles SAM (9 points) pour, de haut en bas, l'évaluation des dimensions de *valence* (du déplaisir au plaisir), d'*arousal* (de calme à excitation) et de *dominance* (de peu de contrôle à contrôle maximal). La déclinaison de ces échelles en 5 ou 7 points est également possible (voir http://irtel.uni-mannheim.de/pxlab/demos/index_SAM.html).

L'arousal et la valence sont reconnues pour être les dimensions décrivant le mieux une émotion (Lang P. *et al.,* 1993), la dominance étant moins fiable et moins stable pour expliquer un état émotionnel (Bradley et Lang P., 1994). Il existe un lien entre les signaux physiologiques et les dimensions d'arousal et de valence, l'activation du système nerveux autonome étant modifiée lors d'une induction émotionnelle. L'activité électrodermale (AED) est généralement associée à la dimension *arousal* et la fréquence cardiaque (FC) à la dimension *valence*.

En effet, diverses études ont montré une augmentation de l'AED lors d'images (Greenwald, Cook et Lang P., 1989 ; Lang P. *et al*., 1993) ou de séquences audiovisuelles définies par un arousal élevé telles que des films à suspense (Hubert et de Jong-Meyer, 1991) ou des scènes fortement désagréables/stressantes (scènes de subincision, circoncision : Lazarus, Speisman, Mordkoff et Davison, 1962 ; Mordkoff, 1964).

Certains attributs d'un contenu ou de sa restitution sont aussi capables de moduler l'arousal et son expression physiologique. La taille de l'écran (Detenber et Reeves, 1996) ou la distance de visualisation (Lombard, 1995) peuvent, par exemple, avoir un impact significatif sur l'évaluation subjective et l'activité physiologique d'un individu. Par exemple, selon Detenber et Reeves, le niveau d'arousal, subjectivement jugé par les spectateurs, augmente lorsque les images sont présentées sur des écrans de visualisation de grandes tailles (augmentation observée entre un écran de 90"[4] par rapport à un écran de 22"[5]). Reeves, Lang A., Kim et Tatar (1999) ont complété ce résultat en montrant que la taille de l'écran influence l'AED qui augmente avec la taille de l'écran. Simons, Detenber, Roedema et Reiss (1999) et Detenber, Simons et Bennett (1998) ont également montré que la présence de mouvements (séquences vidéo *vs.* images fixes présentées durant 6 s) est associée à une augmentation de l'arousal, subjectivement jugé par le spectateur (à partir des échelles SAM) et à une augmentation de l'AED particulièrement lorsque le contenu est jugé avec un arousal fort. L'ensemble de ces résultats met en avant la spécificité de l'AED à refléter le niveau d'arousal.

Le rythme cardiaque serait plutôt un indicateur de la valence émotionnelle (positive ou négative). Plusieurs études ont montré une augmentation de la FC durant la visualisation de contenus fortement désagréables/stressants (Lazarus *et al.*, 1962 ; Mordkoff, 1964). Ces résultats vont dans le sens de la méta-analyse de Cacioppo *et al*. (2000) réalisée à partir d'une vingtaine d'études portant sur la caractérisation physiologique des émotions discrètes[6] (*joie*, *tristesse*, *colère*, *peur*, *dégoût* et *surprise*, Ekman, 1999). Cette méta-analyse a permis d'observer une activation significativement plus importante du rythme cardiaque lors d'un état émotionnel négatif (peur, colère, dégoût et tristesse) par rapport à un état de valence positive (joie). Cependant, cette activation n'est pas toujours constatée. Simons *et al.* (1999) et Detenber *et al.* (1998), lors de leurs études sur l'influence du mouvement, ont observé que le rythme cardiaque permettait de distinguer les émotions positives, neutres et négatives indépendamment de la présence du mouvement. Cependant, comme l'indique la Figure 4.2 ci-dessous (issue de Simons *et al.*), le rythme cardiaque décélérait lors de stimuli déplaisants par rapport aux stimuli plaisants. Cette décélération cardiaque en présence de stimuli déplaisants a également été confirmée lors de la présentation d'images émotionnellement connotées (Winton, Putman et Krauss, 1984). La contradiction observée concernant les résultats sur l'activité cardiaque (augmentation ou diminution) en présence de matériel émotionnellement connoté pourrait s'expliquer par un effet de contenu. Par exemple, Mordkoff (1964) a présenté des scènes de subincision et de circoncision lors de rituels de

4 228,60 cm

5 55, 88 cm

6 Bien que la définition des émotions ne soit pas consensuelle et que plusieurs index de catégorisation coexistent, six émotions discrètes sont toutefois reconnues pour être universelles et supporter les émotions dérivées d'une ou plusieurs de ces émotions principales. Ces émotions dites de base ont été définies par Ekman (1999) et correspondent aux émotions de joie, tristesse, colère, peur, dégoût et surprise.

tribus aborigènes. Ces scènes auraient pu être beaucoup plus aversives que les extraits de Simons *et al.* et Detenber *et al.* (1998) issus de contextes télévisuels et conduire à une réponse défensive de l'organisme traduite par une augmentation de la FC[7].

Par ailleurs, la présence du mouvement a également été à l'origine d'une décélération phasique (court terme) du rythme cardiaque interprétée par les auteurs (Detenber *et al.* ; Simons *et al.*) comme le résultat d'un phénomène attentionnel (le traitement du mouvement nécessiterait plus de ressources attentionnelles).

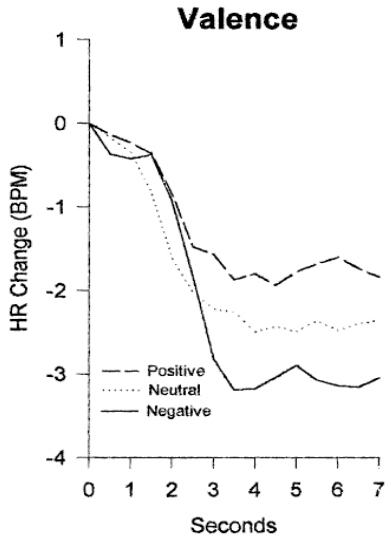


Fig. 4.2. Réponse de la fréquence cardiaque en fonction de la valence (issue de Simons *et al.*, 1999). Le rythme cardiaque est plus élevé pour les émotions positives par rapport aux émotions négatives.

Certains indices de l'activité oculaire seraient également influencés par les états émotionnels d'un individu. C'est notamment le cas pour le diamètre pupillaire et les clignements de l'œil (EB). Selon Chapman, L.J, Chapman, J.P. et Brelje (1969), la dilation pupillaire pourrait refléter un intérêt général ou une évaluation positive ainsi qu'un intérêt sexuel pour un stimulus ou une situation donnée. A l'inverse, une constriction pupillaire est constatée lors de la présentation de stimuli négatifs (Hess, 1972 ; Partala, Jokiniemi et Surakka, 2000). Plus précisément, une dilatation de la pupille serait observée avant la phase de constriction. Hess postule que la réaction émotionnelle produirait une réponse du système nerveux sympathique (dilation pupillaire) et que la constriction surviendrait ensuite, reflétant un état aversif ou d'évitement. Selon lui, un continuum existerait dans lequel la réponse pupillaire à un stimulus serait comprise entre une extrême dilatation, pour un stimulus plaisant ou intéressant et une extrême constriction, pour un stimulus déplaisant ou désagréable. Ce postulat est toutefois controversé (Libby, Lacey, B. et Lacey, J., 1973 ; Woodmansee, 1967). Par exemple, Woodmansee (1967) a observé une dilatation pupillaire pour des images présentant des scènes fortement négatives (scènes de meurtre). Les variations pupillaires

[7] La réaction de défense correspond à un réflexe physiologique de protection, par la préparation de l'organisme à l'action notamment par augmentation de la FC, contre les dangers éventuels sous-tendus par des stimuli intenses, douloureux ou effrayants.

refléteraient également le niveau d'arousal. Une augmentation du diamètre pupillaire a, en effet, été observée lors de la présentation d'images labellisées par un arousal élevé par rapport à des stimuli d'arousal faible (Bradley, Miccoli, Escrig et Lang P., 2008). En plus de la valence et de l'arousal, une diminution du diamètre pupillaire a également été observée lors d'un état d'ennui (Desnoyers, 1987, cité par Gosselin, 2003).

Selon Tecce (1992), une augmentation de la fréquence des clignements de l'œil (EBfreq) reflèterait une expérience émotionnelle négative comme la nervosité ou le stress. A l'inverse, une expérience émotionnelle positive, comme un état de relaxation ou de réussite après une tâche de résolution de problème, serait à l'origine d'une diminution de EBfreq. Tecce émet alors l'hypothèse que l'augmentation des clignements est associée à des sentiments négatifs tandis qu'une diminution serait reliée à une expérience émotionnelle positive.

De manière générale, l'activité physiologique et oculaire semble répondre de manière distincte à la valence d'une émotion ou à son niveau d'arousal. Différentes études ont mis en avant la spécificité du rythme cardiaque à répondre à l'activité émotionnelle, notamment la valence positive ou négative d'un évènement tandis que l'activité électrodermale serait plutôt sensible aux variations d'arousal. L'AED augmenterait avec le niveau d'arousal tandis qu'une décélération de la FC serait observée lors de stimuli déplaisants. En revanche, lors de la présence de stimuli fortement aversifs (mutilations) ou d'induction d'émotions discrètes négatives (peur, colère, dégout et tristesse), la FC augmenterait traduisant une activation physiologique, réaction probablement défensive de l'organisme. Par ailleurs, une dilation du diamètre pupillaire et une diminution des clignements des yeux (EB) seraient observées lors d'une émotion positive.

Au-delà de l'influence de la teneur émotionnelle d'un contenu audiovisuel, certains attributs comme la présence de mouvements, la taille de l'écran de restitution, *etc.* sont également capables d'agir sur la réponse émotionnelle du spectateur et par conséquent, sa traduction physiologique et/ou oculaire. Notamment, une décélération du rythme cardiaque est constatée en présence de mouvements. Cette décélération est attribuée à un phénomène attentionnel appelée réponse d'orientation.

4.2. MODULATIONS ATTENTIONNELLES

L'attention est un système complexe prenant en charge plusieurs fonctions importantes telles que le maintien d'un état d'alerte minimal, l'orientation vers des évènements sensoriels particuliers ou la sélection de stimuli considérés comme pertinents (Posner, Rueda et Kanske, 2007) afin de construire une représentation consciente des stimuli entrants (Posner et Rothbart, 2004, cité par Posner *et al*., 2007). Les états attentionnels varient du sommeil à une attention accrue pour la réalisation d'une tâche ou suite à des stimulations (Posner *et al.*, 2007), en passant par tous les états de veille fluctuant au cours de la journée. Ces états sont différenciés en matière d'intensité et impliquent des mécanismes de commutation des ressources vers un évènement particulier, on parle alors d'aspects sélectifs de l'attention (Posner *et al.*, 2007). Selon Kahneman (1973), l'attention peut en effet être divisée selon des

aspects sélectifs et intensifs. Posner *et al.* divisent l'attention en trois réseaux distincts : l'alerte ou l'état d'éveil (dimension intensive), la réponse d'orientation (orientation automatique vers une nouvelle stimulation) et l'attention exécutive (impliquée dans l'exécution de tâche et la résolution de situation conflictuelle entre plusieurs réponses possibles) pour les aspects sélectifs. Les deux premières dimensions sont notamment importantes à considérer parce qu'elles reflètent respectivement les niveaux toniques (alerte) et phasiques (orientation) de l'attention. En dehors de la subdivision réalisée par Posner *et al.* (2007), il est également possible de considérer les phénomènes d'attention soutenue (Lussier et Flessas, 2001). Lors de visualisation de contenu audiovisuels (télévisuels, cinématographiques) les processus d'alerte, d'attention soutenue (attention portée au contenu) et de réponse d'orientation sont impliqués. Ces différents aspects de l'attention sont présentés dans les paragraphes suivants.

4.2.1. PHENOMENES ATTENTIONNELS TONIQUES

L'*alerte* (parfois appelé *éveil*) correspond à un état énergétique minimal de l'organisme pour le maintien d'une forme d'alarme ou de vigilance générale. Cela doit permettre au système nerveux central d'être réceptif, de manière non spécifique, à toute information endogène ou exogène (Chanquoy, Tricot et Sweller, 2007). L'alerte est souvent associée au concept d'activation physiologique (Caldwell *et al.*, 1994, cité par Clarion) notamment reflétée par une augmentation tonique de l'AED. Plus précisément, la mobilisation énergétique liée à la tâche serait représentée par une augmentation de l'AED par rapport au niveau tonique de repos (Barry *et al.*, 2005). Ainsi, la vigilance ou l'attention soutenue repose probablement, au moins en partie, sur les modifications toniques de l'alerte (Posner *et al.*, 2007).

L'*attention soutenue* peut être définie comme la capacité d'un individu à maintenir, de façon volontaire, son attention sur une tâche particulière de manière ininterrompue ou non perturbée par d'autres stimuli (Chanquoy *et al.*, 2007). Elle est notamment reflétée par la mesure tonique du rythme cardiaque qui montrerait une décélération dans la mesure où cet effort d'attention n'est pas trop important (Lang A., Newhagen et Reeves, 1996 ; Simons *et al.*, 1999). De manière générale, le rythme cardiaque aurait tendance à décélérer lorsque l'attention est portée sur un stimulus extérieur (Lacey, J. et Lacey, B., 1970 ; Lang A., 1990 ; Porges, 1995). Cette décélération serait observée en l'absence de tâche explicite (Berntson, Cacioppo et Fieldstone, 1996) comme cela peut être le cas lors d'une activité de visualisation de contenus audiovisuels. Cette diminution prolongée du rythme cardiaque, durant une attention volontaire, est généralement considérée comme le reflet d'une activation parasympathique (Ravaja, 2004). Durant la visualisation de contenu audiovisuel, l'attention est portée sur un stimulus extérieur (message audiovisuel) sans tâche additive explicite. Ainsi, il est attendu que la FC diminue lors de la visualisation de séquences audiovisuelles (Ravaja, 2004).

Un état d'attention soutenue (Davies et Parasuraman, 1982) et plus généralement les processus de traitement de l'information (voir Stern R. *et al*., 2001) s'accompagnerait également d'une dilatation pupillaire. Par ailleurs, lorsqu'une tâche requiert une attention soutenue, une diminution de la fréquence des EBs (EBfreq) serait observée (Andreassi, 2007). Tecce (1992) a constaté une diminution de EBfreq lorsqu'une attention particulière est portée sur des évènements visuels, peut-être pour faciliter le traitement des informations. De manière générale, une diminution de la fréquence des EBs serait associée au processus d'extraction d'informations d'un environnement visuel (Kramer, 1991). Bauer, Strock, Goldstein, Stern J. et Walrath (1985) explique ce constat par une augmentation de la demande cognitive pour orienter l'attention sur les informations pertinentes d'une stimulation.

4.2.2. PHENOMENES ATTENTIONNELS PHASIQUES : REPONSE D'ORIENTATION

La réponse d'orientation (RO) est un phénomène phasique et un des concepts centraux de la psychophysiologie et notamment de l'étude des médias au travers de ce type de mesures (Ravaja, 2004).

La RO dépendrait du niveau d'éveil physiologique (Barry *et al.*, 2005) et est définie comme une réponse à la fois physiologique et comportementale involontaire dont la fonction est d'orienter l'organisme vers une nouvelle stimulation afin de pouvoir en analyser le contenu et la signification (Boucsein, 2012). La RO est donc exprimée par un ensemble de réponses physiologiques (réponses phasiques) et comportementales incluant : une orientation des organes sensoriels (œil, oreille, *etc*.) en direction du stimulus, une dilatation pupillaire, une vasoconstriction périphérique, une diminution du rythme cardiaque ainsi qu'une augmentation de la conductance électrodermale (Boucsein, 2012 ; Stern R. *et al.*, 2001). La RO peut survenir lors de la présentation d'un stimulus ou de son arrêt ainsi que lors de variation de l'intensité du stimulus[8]. Ainsi, il y a des cas où une mesure physiologique augmente (par exemple, l'AED lors d'une RO) tandis que l'autre diminue (par exemple, la FC[9]) simultanément. Lacey, J. (1967) parle alors de « fractionnement directionnel » de la

[8] Par exemple, Turpin, Schaefer et Boucsein (1999) ont montré une augmentation de l'amplitude et de la fréquence des réponses électrodermales phasiques (REDs) lorsque l'intensité du stimulus augmentait (son de 100 dB par rapport à un son de 60 dB). Ils ont également constaté une accélération cardiaque lors d'une augmentation de l'intensité. Cette accélération ne peut être interprétée comme une RO mais plutôt comme une réponse dite de *sursaut* ou une réponse dite de *défense*. La première est produite en réaction à un stimulus intense et soudain (comme un coup de feu) conduisant au désengagement de l'attention sur l'action en cours, cette réaction se traduit notamment par une accélération immédiate de la FC. La réaction de défense correspond au réflexe physiologique de protection. Dans leur étude, Turpin *et al.* ont toutefois constaté que l'amplitude des REDs était plus sensible à l'intensité que la réponse cardiaque. L'influence de ce paramètre d'intensité est également retrouvée au niveau de l'influence émotionnelle où les variations d'arousal sont mieux reflétées par l'AED.

[9] Il est généralement admis que la décélération du rythme cardiaque, médiatisée par le SNP, reflète le traitement du stimulus (Andreassi, 2007). Cette modification du rythme cardiaque pourrait favoriser ou inhiber les

réponse physiologique. Ce fractionnement tend à démontrer que le principe d'activation ne se réduit pas à un phénomène unidimensionnel (variant d'une faible activation à une forte activation) mais correspondrait plutôt à un phénomène multidimensionnel complexe résultant de l'activité de plusieurs mécanismes énergétiques. A une tâche donnée, mentale ou physique, correspondrait alors un état d'activation particulier (inscrit dans un espace multidimensionnel) pour lequel les performances seraient optimales (Hancock, 1986, cité par Backs et Boucsein, 2000).

La RO est donc une réaction à la nouveauté et de ce fait, une habituation rapide est observée (diminution ou disparition de la RO) après plusieurs répétitions du stimulus. L'activité électrodermale est particulièrement sensible au phénomène d'habituation (Barry, 2004, cité par Clarion, 2009 ; Barry et Sokolov, 1993).

Ces phénomènes attentionnels sont bien connus des réalisateurs et sont utilisés pour maintenir l'engagement du spectateur dans l'activité de visualisation. En effet, selon Lang A., Zhou, Scharwtz, Bolls et Potter (2000), une manière de maintenir l'attention d'un spectateur sur le contenu visualisé consiste à introduire certaines caractéristiques de contenu telles que des changements de plan ou de scène (*cut*), des effets sonores ou encore des messages textuels. Par exemple, plusieurs études ont constaté la présence de RO lors de l'introduction d'informations textuelles à l'écran (Thorson et Lang A., 1992), de *cuts*, de mouvements ou de changement de son (Detenber *et al.*, 1998 ; Lang A., 1990 ; Lang A., 1991 ; Reeves, Thorson, Rothschild, McDonald, Hirsch et Goldstein, 1985 ; Simons *et al.*, 1999) ou de changement de plans (Lang A. *et al.*, 2000). Selon ces auteurs, l'apparition de ces informations engendrerait une RO requérant une augmentation des ressources attentionnelles pour décoder et interpréter le message. Plus le nombre de RO augmenterait, plus les spectateurs prêteraient attention au contenu. Lang A. *et al.* (2000) ont d'ailleurs observé que la conductance cutanée augmente avec le nombre de changement de plan comme l'illustre la Figure 4.3 ci-dessous. De la même façon, d'autres études ont montré que le nombre de changements de plan dans une même scène ou de changements de scène (*cut*) conduit à une augmentation du niveau d'arousal subjectivement reporté et telle que mesurée par la conductance cutanée (Lang A., Bolls, Potter et Kawahara, 1999 ; Yoon, Bolls et Lang A., 1998).

---

traitements corticaux afin de faciliter ou améliorer les entrées sensorielles. Obrist (1981) souligne que cette décélération est accompagnée par une diminution de l'activité motrice : les modulations de la FC seraient un effet indirect de l'attention causé par un apaisement de l'activité motrice. Jennings *et al.* (1992) explique la diminution phasique de la FC par une inhibition de la réponse motrice. Selon lui, la décélération cardiaque doit permettre un traitement plus efficace des stimuli d'entrées grâce à l'inhibition de l'activité motrice en cours. La décélération cardiaque serait alors une sorte de réinitialisation nécessaire pour l'émergence de nouveaux comportements et la réalisation d'opérations basiques et appropriées pour le traitement de l'information. Une décélération cardiaque indiquerait l'initiation et la fin d'un état attentionnel (Andreassi, 2007). Une accélération cardiaque s'ensuivrait manifestant l'allocation des capacités de traitement pour la tâche en cours.

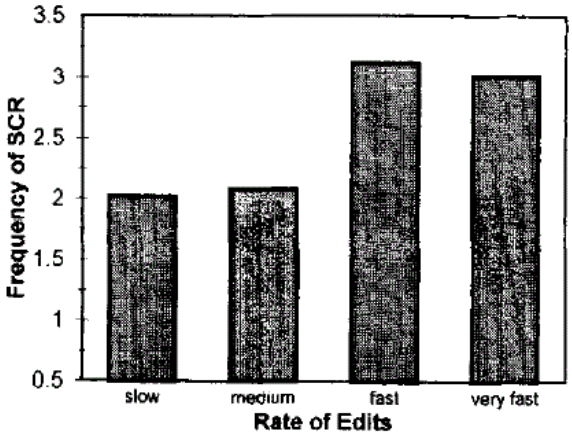


Fig. 4.3. Fréquence de la réponse électrodermale en fonction du nombre de changement de plan où, pour un segment d'une minute, un nombre compris entre 0-7 correspond au niveau « slow », entre 8-15 au niveau « medium », entre 16-23 au niveau « fast », un nombre > 24 correspond au niveau « very fast ».

Hubert et de Jong-Meyer (1991) ont également interprété comme une réponse d'orientation la nette augmentation de l'AED et la décélération temporaire de la FC constatées durant la première minute de chaque contenu de test présenté (extraits audiovisuels de films de 10 min).

4.2.3. APPROCHE A CAPACITE LIMITEE

Selon diverses études, l'attention n'est pas constante durant la visualisation d'un contenu audiovisuel (voir Lang A. *et al*., 2000). L'approche à capacité limitée suppose que les ressources de traitement de l'information du spectateur sont limitées (Lang A., 1995 ; Lang A. et Basil, 1998 ; Lang A. *et al.*, 1999, Lang A. *et al.*, 2000). Selon cette approche, le traitement des messages télévisuels inclut trois étapes : l'étape d'encodage de l'information du contenu, la mise en relation de cette information avec celles déjà stockées en mémoire afin d'en extraire du sens et le stockage de la nouvelle information en mémoire. Ces trois sous-processus se produisent de manière simultanée et continue durant la visualisation. Les ressources attentionnelles, limitées, du spectateur sont distribuées sur l'ensemble de ces trois processus. L'activité de visualisation se réalise de manière optimale lorsque ces trois sous-processus disposent de ressources attentionnelles suffisantes. Dans le cas contraire, l'activité de visualisation est réalisée de manière moins efficace. Quand l'accès à un message est difficile du point de vue de son contenu (message) ou de la structure de contenu (changement de plan, mouvement caméra, *etc.*) voire de leur interaction, les besoins de ressources attentionnelles à allouer à l'encodage ou à l'extraction du sens augmentent. Les ressources attentionnelles supplémentaires allouées aux processus d'encodage et d'extraction du sens pourraient alors être à l'origine d'un effort supplémentaire (allocation énergétique) observable à travers les mesures psychophysiologiques.

4.3. EFFORT MENTAL

Les précédents paragraphes ont indiqué que les mesures psychophysiologiques peuvent être utilisées comme des indicateurs objectifs des émotions humaines ou des phénomènes attentionnels. Dans le domaine de la recherche en facteurs humains, elles sont utilisées pour déterminer l'effort mental (Vicente *et al*., 1987), c'est-à-dire le coût induit par la réalisation d'une activité donnée.

4.3.1. CONCEPT

La notion d'effort mental est intimement liée à la notion de charge mentale au point que leur distinction n'est pas toujours évidente voire mal définie. La notion de charge mentale est largement utilisée dans le domaine de la recherche appliquée aux facteurs humains (opérateurs investis dans une tâche particulière comme la conduite automobile, le contrôle aérien, le pilotage d'avion, *etc*.). Cependant, elle reste une notion complexe et difficile à définir (pour une revue voir Cain, 2007).

La charge mentale est issue de la psychologie du travail et peut être définie comme la charge induite par la réalisation d'une tâche dans des conditions environnementales et opérationnelles spécifiques (Cain, 2007), lorsque les aspects sensorimoteurs de l'activité ne sont pas primordiaux (sinon le terme de charge de travail est privilégié). Elle fait également référence à la capacité maximale de traitement de l'opérateur (Averty, 1998, cité par Clarion, 2009 ; Kahneman, 1973). Clarion précise que toute tâche mentale met en jeu des processus perceptifs qui représentent un coût énergétique pour l'organisme. Enfin, la notion de charge (subie, liée à la tâche) peut être distinguée de la notion d'effort (également liée à la tâche mais non subie car relative à l'action du participant, à ses caractéristiques et à son environnement). Pour Olive, Piolat et Roussey (1997), la notion d'effort renvoie à celle de coût, c'est-à-dire de ressources allouées à un instant donné, au système de traitement pour résoudre une tâche particulière en utilisant une stratégie donnée. L'objectif de la mesure de la charge mentale est de pouvoir quantifier le coût mental lié à la réalisation d'une tâche pour prédire les performances de l'opérateur (Cain, 2007), l'objectif final étant d'améliorer les conditions de travail d'un opérateur, ainsi que les procédures ou interfaces utilisées. La charge mentale peut également être mesurée dans le cadre de l'évaluation de nouvelles interfaces, l'objectif du concepteur étant d'optimiser les performances du système. La charge mentale est en effet considérée comme l'un des facteurs à prendre en considération dans le processus d'optimisation de système (Mitchell, 2000, cité par Cain).

L'effort/charge induit(e) serait également tributaire d'autres facteurs tels que le niveau d'expertise et de développement de l'individu, la stratégie mise en place pour réaliser la tâche (Barouillet, 1996 cité par Chanquoy *et al.*, 2007) ou encore des facteurs relatifs à l'état général de l'individu (Chanquoy *et al.*) comme son seuil de vigilance, son niveau de fatigue, son état émotionnel, *etc*. Par exemple, une émotion négative serait à l'origine d'une focalisation attentionnelle, diminuant la quantité de ressources disponible pour le traitement de la tâche, tandis qu'une émotion positive entraînerait plutôt une ouverture attentionnelle

(prises de décisions complexes plus rapidement, davantage de rejets des mauvaises décisions, évitement de redondances dans les processus de recherche, voir Chanquoy *et al.*).

Les notions d'effort et de charge mentale sont aujourd'hui controversées et de nombreux auteurs travaillent encore à leurs définitions. Toutefois, ces notions font référence aux ressources énergétiques engagées, bien que ne pouvant se résumer par ce seul aspect, pour réaliser une tâche donnée. Ainsi, l'effort pourrait être étudié à travers les dépenses énergétiques qu'il nécessite.

4.3.2. EFFORT MENTAL ET ACTIVATION PHYSIOLOGIQUE

Plusieurs auteurs ont proposé des modèles pour décrire les interactions entre processus cognitifs et allocations énergétiques (Boucsein, 1993 ; Hancock, 1986, cité par Backs et Boucsein, 2000 ; Hancock et Meshkati, 1988 ; Hockey *et al.*, 1986, cité par Backs et Boucsein, 2000 ; Pribram et McGuiness, 1975). Selon ces modèles, tout fonctionnement cognitif nécessiterait à la fois des informations spécifiques à traiter et un apport énergétique, appelé effort, capacité ou attention (Kahneman, 1973).

Kahneman (1973) propose le premier modèle intégrant la notion d'effort. Celui-ci repose sur les théories de capacité attentionnelle limitée[10]. Selon Kahneman, la notion d'effort fait référence à un « potentiel énergétique » permettant l'activation des structures de traitement de l'information ainsi que le contrôle de leur fonctionnement. Ces processus énergétiques permettant de réguler l'état de l'organisme (sect. 3.1.1, chap. III) influenceraient indirectement le traitement de l'information (Gaillard, 1993). Sanders (1990) a proposé une extension du modèle de Kahneman dans le cadre du traitement de l'information. Dans ce modèle, le traitement de l'information est sériel, de l'extraction des composantes du stimulus à la réponse motrice. L'éveil serait impliqué dans les premiers stades du traitement de

[10] Existence d'une capacité centrale contenant des ressources attentionnelles limitées (Kahneman, 1973 ; Norman et Bobrow, 1975). Ces ressources seraient allouées entre les diverses activités cognitives les sollicitant. Lorsque l'attention est dirigée vers les informations pertinentes, pour une situation ou une tâche donnée, les ressources attentionnelles allouées seraient suffisantes pour traiter la tâche de manière efficace. En revanche, en situation d'attention partagée, les ressources attentionnelles seraient distribuées entre les différentes sources d'informations à traiter. Si les ressources attentionnelles nécessaires pour la réalisation des tâches informationnelles dépassent les ressources disponibles alors la réalisation des tâches sera plus difficile (chute des performances) voire impossible pour l'une ou l'autre des tâches. D'autres théories remettent en cause l'existence d'un ensemble unique de ressources attentionnelles et suggèrent plutôt des mécanismes de traitement spécifiques, chacun d'entre eux ayant une capacité limitée, en fonction de la nature de la tâche (Allport, 1980 ; Baddeley, 1986). Cette théorie permettrait d'expliquer la baisse de performances lorsque les tâches sont proches ou similaires. Ces dernières solliciteraient les mêmes mécanismes (c.-à-d. partage des mêmes ressources) tandis que des tâches dissemblables seraient prises en charge par des systèmes de traitement indépendants. L'attention sélective et partagée constituent la base de la théorie de la charge cognitive (Chanquoy *et al.*, 2007).

l'information, c'est-à-dire l'extraction d'indices pertinents par rapport à l'objectif de la tâche, l'activation permettrait de rendre possible ce traitement grâce à une mobilisation énergétique suffisante. L'effort, correspondant à la mobilisation des ressources en attention, résulterait de l'interaction entre éveil et activation. Sanders suppose l'impact de variables computationnelles pouvant moduler la difficulté de la tâche et affecter directement les étapes du traitement de l'information. Notamment, la qualité du signal ou son intensité sont considérées comme des variables computationnelles. Ainsi, l'effort en lien avec une tâche déterminerait le niveau d'activation nécessaire pour sa réalisation.

Selon Collet *et al.* (2003, cité par Clarion) la notion « biologique » d'activation serait assez proche de la notion « psychologique » de charge mentale. Clarion (2009) précise que l'activation varie selon l'attention (dimension intensive) et la vigilance. Ces deux dernières présentent une augmentation lorsque des ressources sont mobilisées par l'organisme pour s'adapter à la demande de la tâche. Selon Clarion (2009), l'activation représenterait un indicateur fiable de la charge mentale.

L'étude du rapport entre la demande (tâche) et les ressources (effort) peut être réalisée selon deux approches à savoir l'astreinte et la contrainte (Gaillard et Kramer, 2000, cité par Clarion, 2009). L'*astreinte* concerne les stratégies mises en œuvre pour réduire le coût de la tâche (sélection des informations pertinentes par exemple) ou encore l'augmentation des ressources par un effort supplémentaire. La *contrainte* concernerait les réactions affectives de l'individu (émotions négatives par exemple). Ces deux approches sont censées entraîner des réactions physiologiques (Clarion, 2009). L'étude de la contrainte permet notamment l'observation de l'effet de la charge de travail des opérateurs sur le long terme (effets psychosomatique, risques sanitaires) (Gaillard et Kramer, 2000, cité par Clarion). L'étude des réactions liées à l'astreinte consistent à observer les changements des dépenses énergétiques consécutives à un effort mental ou à la difficulté de la tâche (dimension intensive, Clarion, 2009). Dans ce cas, les mesures psychophysiologiques peuvent être utilisées comme indicateur de l'effort fourni. De cette manière, l'activation physiologique représente un indicateur de l'effort mental.

Dans ce document, la notion d'effort mental sera réduite à la seule notion d'astreinte, à savoir les dépenses énergétiques (modification du pattern d'activité physiologique, aspect quantitatif) engagées par la tâche de visualisation/écoute de contenus audiovisuels. Par exemple, lorsque l'accès au message audio et/ou vidéo est rendu difficile en raison de la présence de dégradations, des ressources énergétiques supplémentaires seraient engagées pour décoder le signal dégradé. Cette augmentation des ressources (dépenses énergétiques supplémentaires) pourrait être observable principalement à travers les indicateurs de système nerveux sympathique.

4.3.3. EFFORT MENTAL ET MESURES PHYSIOLOGIQUES ET OCULAIRES

Différents auteurs ont montré que des modifications du niveau d'activation du SNA sont souvent associées à des variations de la charge mentale (pour une revue voir Kramer, 1991 ; voir aussi, Grossman *et al.*, 1990 ; Porges, 1992 ; Wilson G. F. et Eggemeier, 1991). En effet, l'AED est reconnu comme un indicateur fiable de l'activation sympathique, par conséquent, elle est considérée comme une mesure valide de l'effort mental (dimension intensive). Par exemple, Kahneman *et al.* (1969), ont constaté une augmentation de l'AED avec le niveau de difficulté d'une tâche mentale (addition de chiffre). De nombreuses études ont confirmé la relation entre AED et effort mental (voir Kramer, 1991 ; voir aussi Gendolla et Krüsken, 2001 ; Ikehara et Crosby, 2005 ; Shi, Ruiz, Taib, Choi et Chen, 2007 ; Wilson G. F., 2002). Plus spécifiquement, l'étude du niveau tonique de l'AED a largement été utilisée comme indicateur de la charge mentale dans des contextes aussi différents que la conduite automobile (Clarion, 2009), le sport (VaezMousavi, Hashemi-Masoumi et Jalali, 2008, cité par Clarion) ou encore chez l'opérateur (Collet, Averty, et Dittmar, 2009, cité par Clarion).

La mesure de la fréquence cardiaque est également reconnue comme un indicateur valide de l'activation physiologique en réaction aux besoins induits par la réalisation d'une tâche. De nombreuses études ont mis en avant une augmentation de la FC avec l'effort mental (pour une revue voir Kramer, 1991 ; Backs et Boucsein, 2000 ; voir aussi Kahneman *et al.,* 1969 ; Porges et Byrne, 1992 ; Roscoe, 1992 ; Wilson G. F., 2002 ; Wilson G. F. et O'donnel, 1988). L'étude fréquentielle de la variabilité du rythme cardiaque et notamment sa composante en basses fréquences permettrait également de refléter les variations de charge mentale. De manière surprenante, il a été constaté que la VRC en général et les basses fréquences en particulier diminueraient lors d'une augmentation de l'effort investi dans la tâche (Backs et Boucsein, 2000 ; Kalsbeek, 1971 ; Kramer, 1991 ; Tattersall et Hockey, 1995 ; Wilson G. F. et O'donnel, 1988). Mulder (1980) a interprété cette réduction comme l'expression d'un pattern émotionnel phylogénétique de défense.

La taille de la pupille varie également en fonction du coût de l'activité mentale. Hess et Polt (1964) furent les premiers à étudier ce lien. Lors d'une tâche de calcul mental (multiplications de 7×8 à 16×23), une augmentation très claire de la taille de la pupille lorsque le niveau de difficulté augmente a été observée. La Figure 4.4 présente les résultats obtenus. La taille de la pupille diminuait immédiatement une fois la réponse donnée. Ainsi, la réponse pupillaire semble refléter la charge liée au processus de traitement de l'information. Plusieurs études, plus récentes, confirment l'augmentation du diamètre pupillaire lors d'un effort mental (Geacintov et Peavler, 1974 ; Juris et Velden, 1977 ; Kahneman, 1973 ; Klingner, Kumar, et Hanrahan, 2008 ; Porter, Troscianko, et Gilchrist, 2007).

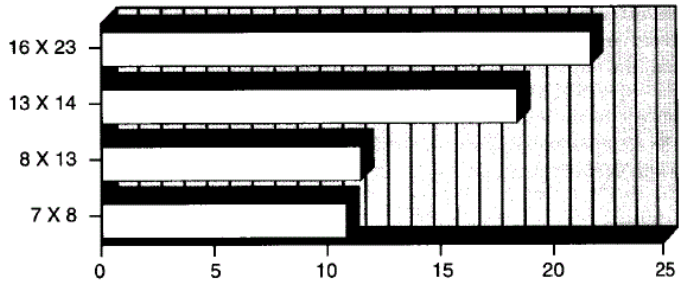


Fig. 4.4. Dilation pupillaire durant la réalisation de calcul mental (Hess et Polt, 1964, figure issue de Beatty et Lucero-Wagoner, 2000). L'ordonnée représente les opérations à réaliser mentalement, l'abscisse présente le pourcentage de dilation pupillaire.

Une autre catégorie de mesure pour observer l'effort mental consiste à étudier les mesures oculaires (Kramer, 1991 ; Wilson G. F. et O'donnel, 1988). Plusieurs études ont montré que la fréquence et la durée des clignements diminueraient avec le nombre d'informations à traiter (Ahlstrom et Friedman-Berg, 2006 ; Veltman et Gaillard, 1998). Par exemple, la fréquence de clignements peut chuter à trois par minute lors d'une activité de lecture (Andreassi, 2007). Une explication consiste à croire que la diminution des EBs lorsque la demande de la tâche augmente pourrait être attribuée à un effort de minimisation du risque de perte d'information importante (Baumstimler et Parrot, 1971), le clignement entraînant une perte temporaire des informations visuelles. La durée des EBs est également associée à l'effort mental lié à une tâche. La durée de fermeture diminuerait avec l'effort mental et augmenterait avec la fatigue (voir Kramer, 1991). De manière générale, les mouvements oculaires (incluant les saccades) auraient tendance à diminuer lors d'un effort mental (May, Kennedy, Williams, Dunlap et Brannan, 1990).

4.4. FATIGUE

La traduction physiologique de la fatigue est généralement attribuée à une domination du système nerveux parasympathique (impliqué dans les états d'hypoactivité). Cependant, certains indices du comportement oculaire comme le PERCLOS ont été spécifiquement élaborés pour étudier le phénomène de fatigue.

La fatigue est un terme difficile à définir. La fatigue physique peut être décrite comme immédiatement consécutive à un effort physique et exprimée par une diminution des performances en lien avec une activité physique (Stern J., Boyer et Shroeder, 1994). La fatigue mentale correspondrait à une diminution des performances liées à une tâche nécessitant un certain niveau de vigilance ainsi que la manipulation et la récupération d'informations stockées en mémoire (Stern J., Boyer et Shroeder, 1994). Plus généralement, la fatigue peut être définie comme « *un ensemble de manifestations engendrées par un effort, qu'il soit intense ou prolongé, ou bien à la fois intense et prolongé* » (Laboratoire d'anthropologie appliquée, 1996).

Selon Stern J., Boyer et Shroeder, la fatigue serait mesurable par des indicateurs physiologiques. Par exemple, plusieurs auteurs ont constaté une influence du niveau de

fatigue sur les variations pupillaires. Plus précisément, le diamètre pupillaire diminue à mesure que la fatigue augmente (Hess, 1972 ; Kahneman et Peavler ; 1969 ; Lowenstein et Loewenfeld, 1964). Cependant, les indices du comportement oculaires (clignements de paupières, par exemple) sont les indicateurs privilégiés de fatigue mentale.

4.4.1. PERCLOS

Dans les années 1980, plusieurs études ont montré que la fatigue pouvait être décrite par divers indicateurs tels que le diamètre de la pupille, la rotation des globes oculaires, le balayage oculaire, le clignement de paupières, *etc.* (Lang L. et Qi, 2008). Cependant, le PERCLOS s'est avéré être l'indice le plus fiable pour détecter la fatigue. La mesure du PERCLOS a été développée et validée pour permettre l'observation, en temps réel, de la baisse de vigilance chez le conducteur (Dinges, Mallis, Maislin et Powell, 1998) et plus généralement chez tout opérateur humain. Il est aujourd'hui reconnu comme un indice fiable de fatigue (Lang L. et Qi, 2008, Wierwille *et al.*, 1994). Pour rappel, le PERCLOS correspond à un indice de détection du pourcentage de fermeture lente des paupières pour une fenêtre temporelle donnée (sect. 3.6.5, chap. III), celui-ci augmente avec la fatigue de l'individu.

4.4.2. SACCADES

Il est généralement accepté que les traitements visuels et cognitifs sont réalisés durant les périodes de fixations (Just et Carpenter, 1984). En effet, les fixations traduiraient l'allocation d'attention visuelle pour la mobilisation des ressources cognitives permettant de porter attention à un élément particulier. Les fixations seraient alors plus longues et plus nombreuses. La fréquence des saccades est plutôt reliée à un état de fatigue. Plusieurs études ont en effet rapporté un ralentissement de la vitesse, de la durée et de l'amplitude des saccades (Bahill et Stark, 1975 ; Schmidt *et al.*, 1979 ; Stern J., Boyer, Schroeder) dans le cadre de l'étude de la fatigue mentale. De ce fait, la fréquence des saccades diminuerait en présence de fatigue (Nakayama, Takahashi et Shimizu, 2002).

4.4.3. CLIGNEMENT DE PAUPIERES (EYE BLINK)

Les EBs peuvent également refléter un état de fatigue (Andreassi, 2007 ; Kramer, 1991 ; Stern R. *et al.*, 2001 ; Tecce, 1992). Spécifiquement, une augmentation de la fréquence des EBs est observée au cours d'une tâche de longue durée notamment si elle conduit à de la fatigue ou de l'ennui (Stern J., Boyer et Schroeder, 1994 ; Tecce, 1992). Leur durée augmente aussi avec la fatigue (Kramer, 1991). En effet, Stern J. et Dunham (1990) rapportent qu'une durée de fermeture prolongée durant le clignement est liée à une baisse de vigilance. De manière générale, le système de contrôle oculomoteur serait très sensible à la fatigue, à la baisse d'attention ainsi qu'à l'ennui (Stern J., 1980).

La visualisation prolongée de contenus audiovisuels pourrait entraîner un état de fatigue notamment visuelle chez le spectateur. Ce problème a particulièrement été traité avec

l'avènement de la vidéo 3D. La question de la fatigue oculaire causée par l'affichage 3D est particulièrement étudiée afin de garantir une bonne expérience au spectateur. Plusieurs études (Eui Chul, Hwan et Kang Ryoung, 2010 ; Jae-Hwan, Byoung-Hoon et Deok-Hwan, 2012) se sont intéressées à ce phénomène en mesurant la fréquence de clignement (EBfreq) lors de la présentation de contenus audiovisuels 2D ou 3D. Il était attendu que la fréquence des EBs augmente avec la fatigue visuelle, cette dernière devant être plus importante lors d'un contenu 3D en raison des efforts de vergence et d'accommodation devant être réalisés (Lambooij *et al*., 2009). Comme le présente la Figure. 4.5 ci-dessous, les résultats ont permis de confirmer ce postulat.

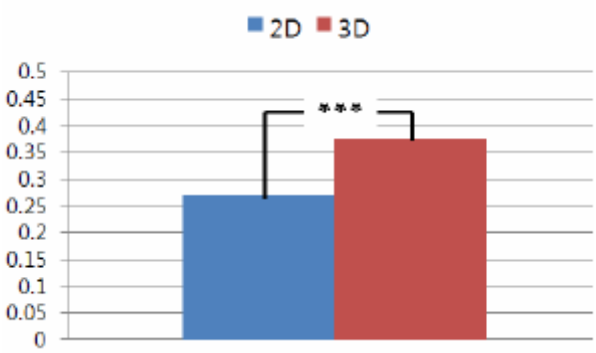


Fig. 4.5. Résultats, issus de l'étude de Eui Chul *et al.* (2010), comparatifs de la fatigue oculaire considérée à travers la moyenne de la fréquence de clignement (EBfreq) entre une visualisation 2D et 3D (* indique un niveau de significativité à p <0,01).**

4.5. VERS UNE MESURE DU COUT UTILISATEUR

Les mesures psychophysiologiques peuvent refléter différents processus : émotionnels, attentionnels ou relatifs au coût d'une activité (effort ou fatigue). Une synthèse des différentes influences des mesures physiologiques et oculaires est proposée par le Tableau 4.1 ci-après.

Les notions d'effort et de fatigue sont particulièrement importantes à considérer dans le cadre de l'optimisation de système de restitution audiovisuelle. En effet, ils pourraient permettre de circonscrire le coût pour le spectateur d'un signal audiovisuel dont la qualité restituée est dégradée et qui pourrait *in fine,* conduire à un rejet du système ou de la technologie utilisée. Une méthode d'évaluation de la qualité audiovisuelle pourrait donc inclure la mesure du *coût utilisateur* en matière de fatigue et d'effort. Les mesures psychophysiologiques se présentent alors comme des candidates pertinentes par leur capacité à refléter ces processus sans être assujetties aux biais relatifs aux mesures subjectives. La combinaison des mesures subjectives et psychophysiologiques au sein d'une même méthode permettrait donc d'étendre l'évaluation de la qualité à la *qualité d'expérience* du spectateur.

Tableau 4.1. Récapitulatif des différentes familles d'influences (émotionnelle, attentionnelle et coût) sur les différents indicateurs physiologiques et oculaires.

		FC	VSP	AED	TCP	DP	EB	PERC-LOS	SAC
EMOTION	VALENCE +	↑	-	-	-	↑	↓ freq	-	-
	VALENCE -	↓/↑	-	-	-	↓	↑ freq	-	-
	AROUSAL +	-	-	↑	-	↑	-	-	-
	AROUSAL -	-	-	↓	-	↓	-	-	-
ATTENTION	ATTENTION (tonique)	↓	-	↑	-	↑	↓ freq	-	-
	RO (phasique)	↓	-	↑	↓	↑	-	-	-
COUT	ACTIVATION (effort mental)	↑ (↓BF/VRC)	↓	↑	↓	↑	↓ dur, freq	-	-
	FATIGUE	-	-	-	-	↓	↑ dur, freq	↑	↓

Une augmentation de la fréquence cardiaque (FC), de l'activité électrodermale (AED), du diamètre pupillaire (DP) et une diminution de la température cutanée périphérique (TCP) correspondent à une activation sympathique, une diminution de la FC et du DP correspondent à une activation parasympathique. La variabilité du rythme cardiaque est notée VRC et sa composante en basse fréquence BF. Les influences des indices oculaires de clignements (EB selon leur durée : EBdur et leur fréquence : EBfreq), de PERCLOS et de saccades (SAC) sont également présentées.

CHAPITRE V – VERS UNE METHODE ALTERNATIVE DE L'EVALUATION DE QUALITE

Les normes actuelles pour l'évaluation de la qualité audiovisuelle sont focalisées sur l'évaluation de la qualité du signal par des participants. Pourtant, la présence de dégradations audio, vidéo ou audiovisuelles pourrait influencer non seulement la perception de qualité jugée au moyen d'une note mais aussi la *qualité d'expérience* (QoE) générale du spectateur notamment envisagée sous l'angle de l'effort ou de la fatigue. Comme présenté dans le chapitre précédent, les indicateurs physiologiques et oculaires sont reconnus pour répondre à ce type de phénomènes. Les indicateurs du système nerveux autonome et du comportement oculaire sont donc des candidats pertinents pour permettre l'évaluation de la QoE.

Différentes études ont intégré les mesures psychophysiologiques pour aborder la notion de l'expérience utilisateur à travers le coût de l'utilisation d'une technologie pour l'individu notamment dans le domaine de l'interaction homme-machine tel que le jeu vidéo (Lin, Omata, Hu, Imamiya, 2005 ; Mandryk, Inkpen et Calvert, 2006 ; Ravaja, Saari, Laarni, Kallinen, Salminen, Holopainen, et Järvinen, 2005 ; Wu et Lin, 2011), les environnements virtuels (Meehan, Insko, Whitton et Brooks, 2002) ou les sites internet (Tuch *et al*., 2011 ; Ward et Marsden, 2003 ; Ward, Marsden, Cahill et Johnson, 2002). En revanche, peu d'études se sont attachées à étudier la pertinence des indicateurs psychophysiologiques dans le cadre de la qualité audio et/ou vidéo. L'objectif de ce chapitre est de présenter les quelques études disponibles et qui seront utilisées dans ce document comme référent méthodologique. Une méthode hybride incluant à la fois la mesure de l'expérience subjective et celle de l'activité physiologique et oculaire du spectateur sera proposée comme alternative aux méthodes actuelles d'évaluation de la qualité audiovisuelle.

5.1. EVALUATION DU COUT UTILISATEUR

Le terme *coût utilisateur* (user cost) est emprunté au domaine des interactions homme-machine ou homme-homme médiatisée (visioconférence) pour désigner le niveau d'investissement nécessaire pour atteindre et maintenir un niveau élevé d'utilisabilité envisagée sous l'angle des performances ou de la satisfaction de l'utilisateur par exemple. Le coût pour l'utilisateur peut notamment être étudié sur le plan de l'effort mental ou du stress/anxiété physique et/ou mental (Sweeney, Maguire et Shackel, 1993, cité par Lin, Imamiya, Hu et Omata, 2007).

Pour Wastell et Newman (1996) une compréhension holistique du comportement humain repose sur l'étude conjointe de trois dimensions fondamentales : le comportement manifeste, la physiologie et l'expérience subjective. Pour étayer leur approche, Wastell et Newman ont combiné des indices physiologiques (pression artérielle -systolique et diastolique- et fréquence cardiaque), subjectifs (échelles catégorielles) et de performances, dans l'intention d'évaluer le stress de répartiteurs d'ambulances à la suite du passage de l'utilisation d'un support papier à un système informatique. Leurs résultats ont montré que l'utilisation du

support informatique améliorait les performances (comportement manifeste), diminuait le stress (physiologie) ainsi que l'anxiété subjectivement reportée (expérience subjective). La qualité d'un système a donc été jugée sur la façon dont elle a affecté non seulement les performances mais aussi la *qualité d'expérience* de l'utilisateur. Cette étude a révélé que les mesures psychophysiologiques pouvaient être un outil pertinent d'évaluation de système et plus généralement, que les méthodologies d'évaluation devraient davantage tenir compte de ces aspects.

Ward *et al.* (2002) et Ward et Marsden (2003) ont recueilli les indices d'activité électrodermale (AED), de volume sanguin périphérique (VSP) et de fréquence cardiaque (FC) pour deux groupes d'utilisateurs en situation de navigation internet (durant 10 min). Deux versions d'un même site web était présentées : une version « bien conçue » (respect des règles d'ergonomie) et une version « mal conçue » (présence excessive de menus déroulants - information difficile à trouver-, nombreuses fenêtres publicitaires, déplacements soudains du contenu de la page, peu d'aide à la navigation, *etc.*). Chaque groupe naviguait sur l'une des deux versions. Il était attendu que le coût pour l'utilisateur soit plus élevé (navigation difficile) pour la version « mal conçue » et que cela soit reflété par les données physiologiques des utilisateurs. Des tendances distinctes, obtenues à partir de l'étude de la moyenne des indices recueillis (approche tonique) ont été observées entre les deux groupes. Les utilisateurs de la version « bien conçue » ont eu tendance à se détendre après la première minute, alors que les utilisateurs de la version « mal conçue » ont montré un niveau élevé d'activation, associé à un état de stress et reflété par une augmentation de l'AED et de la FC moyenne, et ce tout au long de la tâche .

Lin *et al.* (2007) ont étudié le *coût utilisateur* de jeux vidéo à travers la variabilité du rythme cardiaque (ou VRC étudiée à travers la bande basse fréquence : BF), le diamètre pupillaire et les mesures subjectives (évaluation de l'effort mental). Trois parties de dix minutes étaient jouées par dix participants. Chaque partie présentait un niveau de difficulté différent : débutant, intermédiaire et expert. Après chaque partie, les participants disposaient de cinq minutes pour se reposer et évaluer l'effort mental ressenti durant la tâche. Les résultats subjectifs ainsi que la VRC ont reflété les niveaux de difficulté du jeu. Une augmentation de la pupille a également été observée entre le niveau débutant et les niveaux intermédiaire et expert. Une corrélation entre les données de VRC et les mesures subjectives a également été observée pour les trois niveaux de difficultés. Ces résultats indiquent que mesures subjectives et physiologiques sont pertinentes pour évaluer les variations du coût pour l'individu de l'utilisation d'une technologie.

Vicente *et al*. (1987) ont également étudié le lien entre mesures physiologiques (VRC à travers l'étude des BF) et le *coût utilisateur* lors d'une tâche psychomotrice (utilisation d'un simulateur d'*hovercraft*). Les résultats ont indiqué que les données physiologiques étaient significativement corrélées aux évaluations subjectives d'effort, mais pas de charge de travail ou de difficulté de la tâche. Dans le domaine de l'interaction homme-machine, d'autres

chercheurs ont également utilisé la VRC comme un indicateur de l'effort mental (Rani, Sims, Brackin et Sarkar, 2002 ; Rowe, Sibert et Irwin, 1998 ; Tattersall et Hockey, 1995).

Tuch *et al.* (2011) ont étudié l'influence du niveau de complexité visuelle (définie selon la taille, en bps, du fichier compressé) de pages Web sur l'utilisateur. Quarante-huit participants ont visualisé un total de trente-six pages Web (copies d'écran de sites existants) de complexité visuelle différente durant huit secondes pendant lesquelles des mesures oculaires (amplitude de la saccade) et cardiovasculaires (fréquence cardiaque) étaient enregistrées. Ces mêmes images étaient présentées une seconde fois pour permettre aux participants de les évaluer à partir des dimensions de valence et d'arousal (échelle SAM, sect. 4.1, chap. IV), les niveaux d'attrait et de gêne étaient également jugés (échelle de Likert). Les pages les moins complexes ont été associées à une valence davantage positive ainsi qu'à une diminution des mouvements oculaires (en particulier dans les premières secondes de visualisation). Une décélération cardiaque a aussi été observée durant la visualisation des images présentées, tous niveaux de complexité confondus. Cette décélération était plus importante pour les images présentant une forte complexité : pour les auteurs, des stimuli complexes entraîneraient une réponse d'orientation plus importante.

Ces études ont montré que des indicateurs tels que l'AED, la FC, les mouvements oculaires ou encore des indicateurs de la VRC sont pertinents pour évaluer le coût pour l'individu lors de l'utilisation de technologies nouvelles ou existantes. Ces résultats montrent que la qualité de l'expérience utilisateur peut être reflétée non seulement par des mesures subjectives mais aussi par des mesures physiologiques et oculaires. Ces dernières peuvent alors s'intégrer aux méthodes pour l'optimisation de la relation entre l'utilisateur et une technologie donnée telle que celle du domaine de la restitution audiovisuelle.

5.2. COUT UTILISATEUR ET EVALUATION DE QUALITE

Wilson G. M. et Sasse (2000a, 2000b) ont proposé de nouvelles méthodes pour l'évaluation de la qualité audio ou vidéo dans le cadre de l'étude du *coût utilisateur* pour des services de communication multimédia (visioconférence). Selon ces auteurs, l'évaluation subjective de la qualité est utile car elle permet de mesurer le degré de satisfaction de l'utilisateur par rapport au niveau de qualité. Cependant, elle ne serait pas nécessairement un indicateur fiable de l'impact de la qualité sur l'utilisateur. Selon Knoche, De Meer et Kirsh (1999), il est impossible pour les individus d'enregistrer, et donc d'évaluer, ce qui n'a pas été consciemment perçu. Sur la base de ce constat, Wilson G. M. et Sasse proposent d'utiliser des mesures psychophysiologiques pour une compréhension holistique de l'influence de la qualité sur l'utilisateur. Wilson G. M. et Sasse formulent l'hypothèse selon laquelle un utilisateur, face à une situation de qualité audio et/ou vidéo insuffisante, doit réaliser un effort perceptif supplémentaire pour décoder les informations du message. Cet effort devrait être reflété par une réponse physiologique (activation du système nerveux sympathique) assimilée à une situation d'inconfort ou de stress même si l'utilisateur est toujours en mesure d'exécuter la

tâche principale. L'activation physiologique est alors considérée comme un évènement négatif. Selon les auteurs, un niveau d'activation de base serait inhérent à la réalisation de la tâche, l'ajout de dégradations de la qualité augmenterait ce niveau.

Plus précisément, Wilson G. M. et Sasse ont proposé de compléter les évaluations subjectives par des mesures physiologiques dans le cadre de l'étude du *coût utilisateur* de services de communication multimédia lors de qualités audio (2000b) ou vidéo (2000a) dégradées. La première étude présentait des séquences audiovisuelles (contexte passif de visioconférence) selon deux niveaux de qualité vidéo, élevé ou faible. La seconde expérience consistait à étudier l'effet de différents niveaux de qualité audio (séquence audio seule) sur l'auditeur. Dans les deux expériences, les indices recueillis correspondaient à l'AED (activité électrodermale), la FC (fréquence cardiaque) et le VSP (volume sanguin périphérique). Les auteurs s'attendaient à une augmentation de la FC et de l'AED ainsi qu'à une diminution du VSP lors des conditions de qualité dégradée. Pour recueillir ces mesures, les auteurs ont utilisé le matériel *Procomp* (Thought Technology Ltd.). La qualité perçue audio (2000b) ou audio et vidéo (2000a) a également été évaluée à partir d'une échelle continue (0 à 100) sans labels. Afin d'éviter toutes contraintes exercées sur les participants par des facteurs autres que la qualité, les auteurs ont fait en sorte que l'environnement de test soit le moins stressant possible (pas de sonneries de téléphone par exemple). Une baseline de quinze minutes était enregistrée pour chaque participant avant l'expérimentation. Cela devait permettre d'une part, aux participants de s'habituer à la présence des capteurs, et d'autre part, de recueillir suffisamment de mesures afin de pouvoir ensuite comparer ces réponses à celles obtenues durant l'expérience.

Comme présenté dans le chapitre III, l'activité physiologique peut être étudiée selon deux approches :

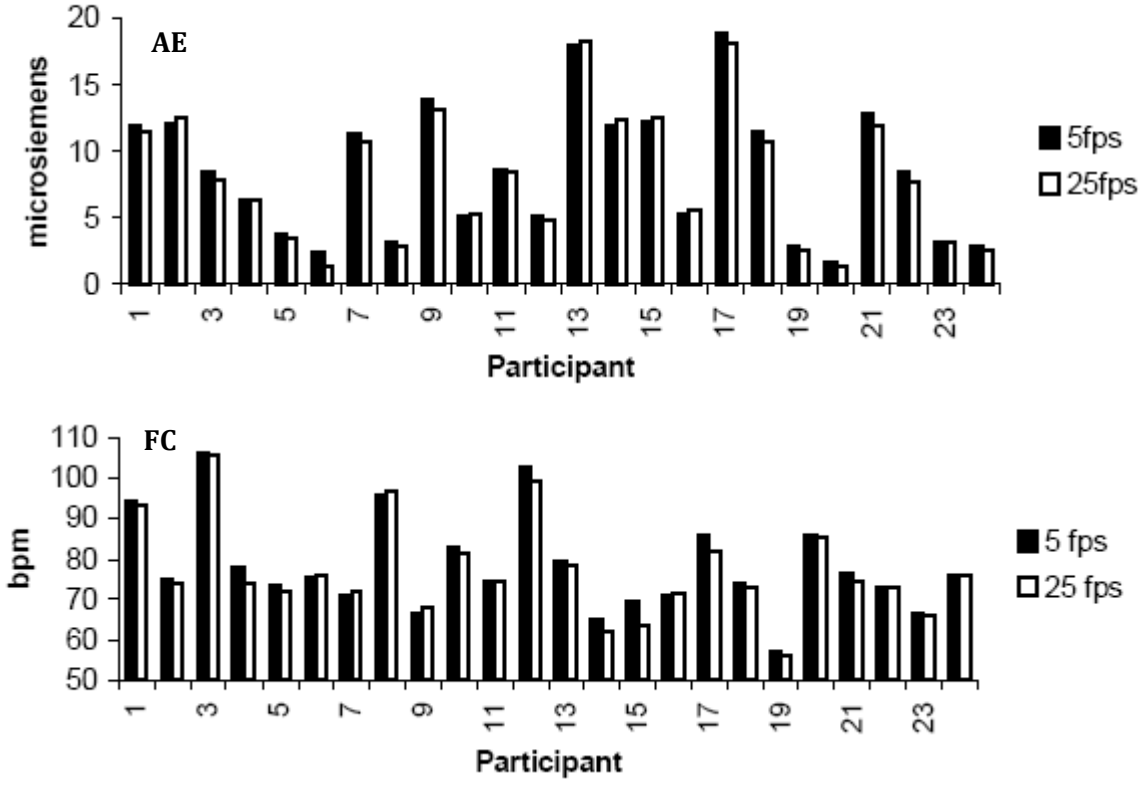
- l'étude de l'activité **physiologique tonique** par la comparaison de valeurs physiologiques (comme la moyenne) obtenues pour des intervalles de temps donnés (conditions expérimentales différentes),
- l'étude de l'activité **physiologique phasique** par la détection de changements physiologiques à court-terme (quelques secondes) en réponse à des évènements spécifiques (Ward et Mardsen, 2003).

Wilson G. M. et Sasse ont utilisé l'approche tonique pour étudier les mesures psychophysiologiques recueillies. Les moyennes, pour chaque signal, obtenues soit pour chaque participant et chaque condition de qualité (2000a) soit pour chaque condition de qualité tous participants confondus (2000b) ont été utilisées.

Dans la première expérience (2000a), vingt-quatre participants devaient visualiser deux entretiens préenregistrés entre deux élèves et un tuteur dans le cadre d'une candidature pour une admission universitaire. Chaque entretien durait quinze minutes. Il était demandé aux participants de porter un jugement sur les candidats à l'issue des entretiens. Durant la

visualisation, la qualité audio était bonne et ne variait jamais. En revanche, la vidéo présentait deux différents niveaux de qualité : élevée (25 images par seconde ou ips) ou faible (5 ips). A titre indicatif, la synchronisation entre parole et image du locuteur est perçue à partir de 16 ips tandis que la perception d'un mouvement fluide est fixée à 25 ips. Durant les cinq premières minutes de l'expérimentation, la vidéo était présentée avec un taux de 16 ips. Ce segment était ensuite exclu du jeu de données pour éviter un effet biaisant lié au passage de l'état de repos à l'expérimentation. Après cette première étape, les entretiens étaient présentés aux participants. Chaque interview présentait un pattern de variation de qualité lui étant propre : 5-25-5 ips ou 25-5-25 ips. Chacune de ces trois périodes durait cinq minutes. Les participants évaluaient la qualité de l'audio/vidéo durant la visualisation (échelle continue) puis la qualité audio et vidéo après la visualisation. Les mesures physiologiques ont été enregistrées tout au long de l'expérience.

Le niveau tonique de chaque signal a donc été étudié à travers la moyenne obtenue pour chaque niveau d'ips (5 ou 25) pour chacun des vint-quatre participants. Les résultats, illustrés par la Figure 5.1 ci-dessous, ont montré une augmentation significative des moyennes d'AED et de FC ainsi qu'une baisse significative de la moyenne du VSP lors des périodes présentées à 5 ips par rapport aux périodes à 25 ips. Ce résultat indique que la qualité vidéo influence l'activité physiologique de l'utilisateur. Ici une activation du système nerveux sympathique est observée. Les auteurs confirment leur hypothèse et concluent qu'un niveau de qualité vidéo faible entraînerait un état de stress.



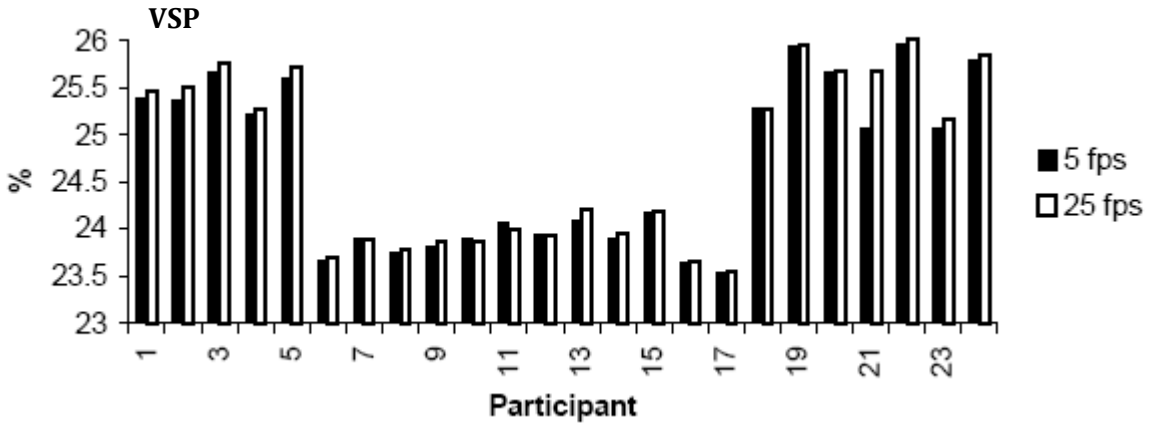


Fig. 5.1. Moyennes obtenues pour chaque participant et chaque niveau de qualité vidéo (5 ou 25 ips) pour chaque indicateur physiologique de haut en bas : AED, FC et VSP.

Par ailleurs, 84% des participants ont déclaré ne pas avoir remarqué de différences de qualité. Aucune corrélation entre les réponses subjectives et physiologiques n'a en effet été observée. Ce constat pourrait être expliqué par un effet de la modalité dominante. Il est en effet possible de supposer que les entretiens présentés apportaient des informations essentiellement auditives et auraient plutôt orienté l'attention du participant vers la modalité audio au détriment de la modalité vidéo dégradée (Hands, 2004). **Wilson G. M. et Sasse concluent que les mesures physiologiques ne sont pas soumises aux mêmes contraintes que les évaluations subjectives et que, lorsque les utilisateurs sont engagés dans une tâche, ils ne remarquent pas toujours la différence entre deux niveaux extrêmes de qualité, pendant ou après la tâche, alors que cette différence est enregistrée au niveau physiologique**. Ainsi, le niveau de qualité a un impact sur l'activité physiologique de l'utilisateur. Ce résultat révèle l'intérêt de l'ajout des mesures psychophysiologiques aux mesures subjectives qui ne sont pas en mesure de rendre compte de la totalité de l'influence de la qualité sur l'utilisateur. Ainsi, les niveaux de qualité, pour une application télévisuelle ou de visioconférence, ne peuvent être définis uniquement sur la base des seules évaluations subjectives. Les concepteurs d'applications et les fournisseurs de réseaux ou de services doivent donc considérer cette information.

La seconde étude de Wilson G. M. et Sasse (2000b) portait sur l'étude de la qualité audio seule. Vingt-quatre participants étaient invités à écouter un dialogue de deux minutes entre deux locuteurs. Celui-ci était présenté six fois ; à chaque présentation, une nouvelle dégradation était introduite : perte de paquets (5% ou 20%), écho, variation du volume de la voix d'un des deux interlocuteurs (faible ou fort) et distorsion (enregistrement d'un des deux locuteurs avec un mauvais micro). Chaque condition était entendue deux fois et une condition référence (non dégradée) était présentée au début et à la fin de l'expérimentation. La qualité audio était évaluée après l'écoute de chaque séquence. Les mesures physiologiques ont été enregistrées tout au long de l'expérience.

Dans cette étude, les niveaux toniques ont été étudiés à partir de la moyenne de chaque indicateur pour chaque condition de qualité obtenue pour l'ensemble des participants. Les résultats ont indiqué un effet significatif de la qualité sur les moyennes de la FC et du VSP, en revanche, les conditions n'ont pas influencé de manière significative l'AED moyenne. Les résultats obtenus sont présentés par la Figure 5.2 ci-dessous.

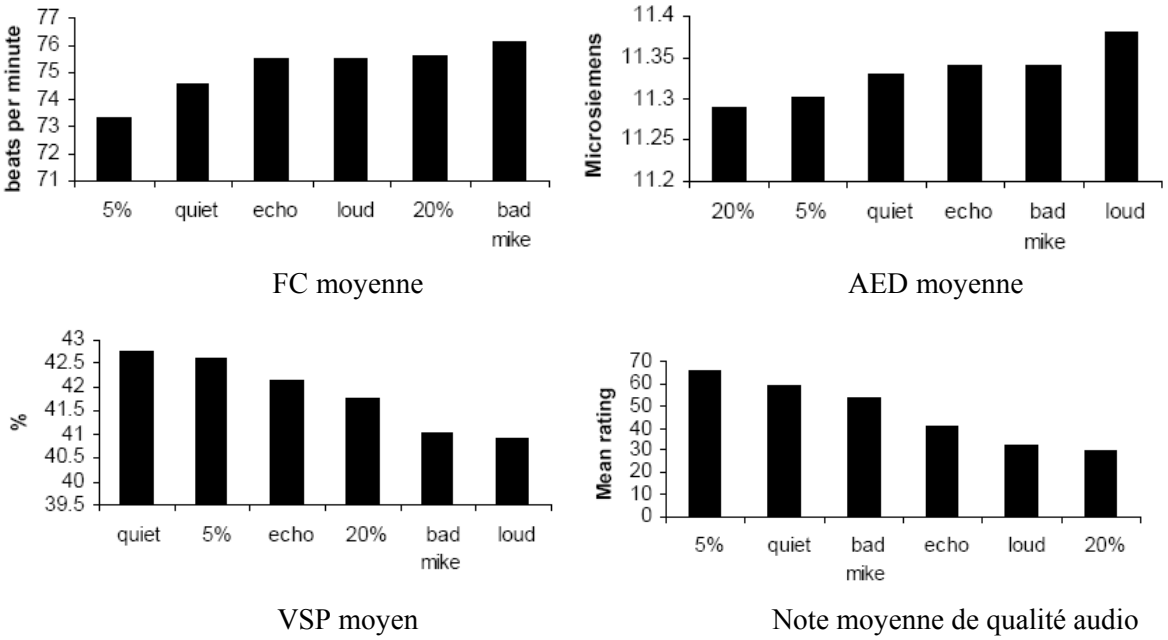


Fig. 5.2. Moyennes obtenues sur l'ensemble des participants pour chaque condition de dégradation (perte de paquets 5 et 20%, écho, variation du volume – quiet ou loud- et distorsion –bad mike-) et chaque mesure physiologique (AED, FC, VSP) et subjective (échelle de 0 à 100).

Comme cela peut être observé à partir des figures ci-dessus, la dégradation de distorsion (*bad mike*, liée à un mauvais enregistrement) correspondait à la dégradation ayant entraîné la plus forte activation physiologique du point de vue de la FC et du VSP (respectivement première et deuxième position). Pour les auteurs, cette dégradation correspondait à la plus stressante du point de vue de l'activité physiologique. L'effet de la dégradation de distorsion n'est en revanche pas reflété par les mesures subjectives. En effet, celles-ci indiquent que la distorsion était classée parmi les conditions dégradant le moins la qualité perçue. Les auteurs expliquent cette différence entre mesures physiologiques et subjectives par un effet de la tâche. Selon eux, la tâche d'écoute, de courte durée (2 minutes par condition), ne serait pas suffisante pour que l'influence de la dégradation de distorsion se manifeste, du point de vue du jugement conscient, sur l'auditeur. En outre, la tâche était passive, donc les effets de distorsion du signal audio peuvent ne pas avoir affectés les participants comme cela aurait pu être le cas en situation d'interaction.

Du point de vue des évaluations subjectives, la dégradation « 20% de perte de paquets » a été perçue comme dégradant le plus fortement la qualité audio. Les auteurs supposent que si cette dégradation n'a pas été à l'origine de la plus forte activation physiologique observée

c'est parce qu'elle était moins « stressante » que la dégradation de distorsion. L'effort pour accéder au sens était moindre lors de la perte de paquets que lors de la distorsion, plus irritante pour les utilisateurs. Le fait que les participants aient jugé la condition « perte de paquets » comme dégradant plus fortement la qualité pourrait provenir du fait que celle-ci dégradait la séquence de manière continue tandis que la distorsion affectait la qualité de manière plus fluctuante. Un effet d'une altération de l'intelligibilité du message pourrait également être supposé pour expliquer les résultats concernant la perte de paquets.

Concernant les autres dégradations, l'augmentation de volume (loud) a été à l'origine d'une perte de qualité subjectivement et physiologiquement reportée. Cette dégradation a plus fortement altérée la qualité que la condition « volume faible » (quiet). La dégradation par présence d'écho a été jugée avec des notes de qualité basses mais ne semble pas être reflétée par les mesures physiologiques (activation plus importante de la FC uniquement par rapport à la dégradation « quiet »). Enfin, la dégradation par perte de paquets à 5% n'a pas été considérée comme un problème aussi bien du point de vue de la mesure subjective que physiologique.

Ces résultats indiquent que les mesures psychophysiologiques et subjectives apportent des informations différentes notamment pour des dégradations de type « perte de paquets » ou « distorsion ». Ces résultats confirment l'intérêt de compléter les mesures subjectives par des mesures psychophysiologiques pour étudier la manière dont un utilisateur est affecté par le niveau de qualité audio et/ou vidéo.

Ces deux expériences permettent plusieurs conclusions. La première, peut-être la plus importante, est que des réponses toniques de l'activité physiologique peuvent être détectées en réaction à la présence de dégradations audio ou vidéo. Un second constat, également important à considérer, concerne le fait que les signaux physiologiques recueillis ne sont pas toujours congruents, certains pourraient répondre plus spécifiquement à certaines dégradations. Par exemple, l'AED a réagi fortement au changement du nombre d'ips mais n'a pas répondu aux dégradations audio alors plutôt reflétées par les variations de FC. Selon Wilson G. M. et Sasse, cela pourrait aussi signifier que la modalité de la dégradation, audio ou vidéo, influencerait différemment l'activité physiologique de l'utilisateur. Ce postulat peut être rapproché du concept d'activation multidimensionnelle supposant qu'une activité donnée correspondrait à un pattern d'activation donné pour lequel les performances seraient optimales (voir sect. 4.2.2, chap. IV). Ainsi, le traitement de séquences audiovisuelles ou de séquences audio seules ou le traitement de dégradations audio ou vidéo conduirait à des patterns d'activation différents et spécifiques.

Enfin, l'évaluation subjective n'est pas toujours corrélée avec les réponses physiologiques. Ce constat est considéré par Wilson G. M. et Sasse comme un argument en faveur de l'intégration de mesures psychophysiologiques pour l'évaluation de qualité de services audiovisuels. Pour les auteurs, ces résultats sont encourageants dans le cadre de l'élaboration d'une méthode reposant sur l'approche tripartite proposée par Wastell et Newman à savoir que les mesures de performances (dans le cadre de l'utilisation d'un système tel que les

communications interactives) et de satisfaction utilisateur (envisagée sous l'angle de la qualité perçue) doivent être utilisées conjointement avec les mesures du *coût utilisateur* (à travers les mesures psychophysiologiques). Wilson G. M. et Sasse concluent qu'en conséquence de ces résultats une plus grande considération devrait alors être apportée à la notion de *coût utilisateur* d'un service lors de l'évaluation du niveau de qualité et plus généralement lors de l'évaluation de l'utilisabilité d'une technologie.

5.3. VERS UNE METHODE HYBRIDE

Comme l'ont révélé les études de Wilson G. M. et Sasse (2000a, 2000b), des mesures toniques de l'activité physiologique sont de bonnes candidates pour l'évaluation de la *qualité d'expérience*. Elles permettent d'obtenir un retour sur le coût, pour l'utilisateur, induit par une qualité audio ou vidéo dégradée. Ces auteurs ont montré que les mesures psychophysiologiques permettent d'observer des effets de la qualité qui ne sont pas toujours captés par les mesures subjectives. Leurs études ont en effet révélé que des fluctuations importantes de qualité peuvent ne pas être consciemment perçues par l'utilisateur mais reflétées par l'activité physiologique. Ce constat indique que l'évaluation subjective mesurée par des notes de qualité perçue seules n'est pas suffisante pour refléter pleinement l'impact de la qualité audiovisuelle restituée sur le ressenti de l'individu. Ces informations sont particulièrement importantes dans une optique d'optimisation de système.

Les travaux de Wilson G. M. et Sasse portaient sur l'évaluation de services interactifs multimédias mais leurs études ont été réalisées pour des contextes passifs (non interactifs) de visualisation et/ou écoute de séquences de test. Leur cadre théorique, à savoir la réalisation d'un effort perceptif pour décoder un signal dégradé, peut donc être transposé à un contexte de visualisation passive (télévisuelle, cinématographique) de contenus audiovisuels.

5.3.1. HYPOTHESES GENERALES

Les chapitres précédents ont indiqué que tout effort, physique ou cognitif, est lié à une dépense énergétique nécessaire à sa réalisation et que les mesures psychophysiologiques permettent de refléter ces variations. Il est envisageable que les variations de qualité du son et/ou de l'image conduisent à des fluctuations des dépenses énergétiques engagées. Une hypothèse consiste à supposer que la présence de dégradations serait à l'origine d'une augmentation de l'état énergétique, par rapport à une condition sans dégradations, c'est-à-dire à un état d'activation physiologique (pouvant être lié à un effort perceptif, une allocation supplémentaire de ressources attentionnelles, *etc.*). Les dépenses énergétiques engagées en présence d'une qualité dégradée seront interprétées comme l'expression d'un effort mental pour décoder le message altéré. L'activation physiologique pourra alors être mesurée à travers différents indicateurs physiologiques et oculaires. Précisément, une domination du système nerveux sympathique (sur le système parasympathique), traduite par une augmentation de la fréquence cardiaque (FC), de l'activité électrodermale (AED), du diamètre pupillaire (DP) et par une diminution du volume sanguin périphérique (VSP) et de la température cutanée

périphérique (TCP), devrait être observée lors d'un segment audiovisuel dégradé (effort mental) par rapport à un segment non dégradé. Des modifications du comportement oculaire, traduites par une diminution de la durée et de la fréquence des clignements de l'œil (EBdur et EBfreq), peuvent aussi être attendues. Ces variations seront recueillies à travers l'étude tonique de l'activité oculaire et physiologique comme proposé par Wilson G. M. et Sasse.

Par ailleurs, le *coût utilisateur* (entendre le coût de l'activité de visualisation pour le spectateur) lorsque la qualité AV fluctue pourrait également être envisagé sous l'angle de la fatigue consécutive à un effort mental répété ou prolongé. Cet effet pourrait être d'autant plus marqué que les dégradations surviendraient sur des contenus 3D (à l'origine de fatigue visuelle, sect. 4.4.3, chap. IV) ou que la séquence visualisée serait longue. La fatigue est généralement étudiée à partir des variations du comportement oculaire (clignement de paupière, fermeture de l'œil, saccades) dont certains indicateurs ont été définis pour rendre compte de manière spécifique des phénomènes de fatigue. Ceux-ci seraient alors traduits par une augmentation du PERCLOS, de EBfreq et EBdur et par une diminution du DP et du nombre de saccades (SAC).

Ainsi, un message audio et/ou vidéo altéré par la présence de dégradations pourrait induire un effort mental supplémentaire pour décoder le signal dégradé puis un état de fatigue observables à travers les indicateurs du système nerveux sympathique et de l'activité oculaire.

L'approche psychophysiologique proposée sera utilisée pour évaluer les changements d'états de l'organisme sous l'influence de l'activité de visualisation de contenus audiovisuels 2D ou 3D présentant des fluctuations de qualité (c.-à-d. avec des dégradations qui ne sont pas toujours équivalentes ou de même nature au sein d'un même contenu de test comme cela pourrait être le cas en conditions réelles de visualisation). Les variations des mesures psychophysiologiques attendues lors d'un effort mental ou d'un état de fatigue en présence de dégradations sont résumées par le Tableau 5.1 ci-dessous.

Tableau 5.1. Effet d'un effort mental ou de fatigue potentiellement induits par la présence de dégradations sur les mesures physiologiques et oculaires. La variabilité du rythme cardiaque est notée VRC et sa composante en basse fréquence BF.

	FC	VSP	AED	TCP	DP	EB	PERCLOS	SAC
EFFORT MENTAL	↑ (↓ BF/ VRC)	↓	↑	↓	↑	↓ dur, freq	-	-
FATIGUE	-		-	-	↓	↑ dur, freq	↑	↓

En-dehors du coût pour le spectateur[11], étudié sous l'angle de l'effort mental et de la fatigue, la qualité audiovisuelle pourrait également influencer l'expérience subjective du

[11] En-dehors de la notion de *coût utilisateur*, le terme de spectateur sera privilégié, pour plus de clarté, dans la suite du document par rapport au terme d'utilisateur en raison de la nature passive de l'activité de visualisation, l'individu n'ayant aucune action sur le système (pas d'interaction homme-système).

spectateur entendue au sens large. Actuellement, les notes subjectives de qualité limitent la mesure de satisfaction, quant à la qualité audiovisuelle restituée, à la seule perception de qualité. Pourtant d'autres aspects comme le niveau de compréhension, d'intérêt, de plaisir, *etc.* pourraient être influencés par des dégradations de qualité et diminuer la *qualité d'expérience* globale du spectateur (QoE). Une seconde hypothèse consiste donc à croire que la qualité audiovisuelle influence non seulement la qualité perçue mais aussi des facteurs subjectifs additifs et déterminants pour la *qualité d'expérience*. Un questionnaire enrichi proposant des mesures subjectives supplémentaires doit être envisagé pour mieux comprendre le lien entre qualité audiovisuelle restituée et *qualité d'expérience*.

Les mesures subjectives et psychophysiologiques pourraient être mises en regard afin d'observer leurs éventuelles influences réciproques. Par exemple, la présence d'effort mental ou de fatigue pourrait diminuer l'expérience subjective en matière de qualité perçue mais aussi d'intérêt, de plaisir, *etc.*

5.3.2. OBJECTIFS

L'objectif de ce travail est de proposer une méthode alternative aux méthodes actuelles d'évaluation de la qualité audiovisuelle dans l'intention d'obtenir un retour plus fidèle et plus complet de l'influence de la qualité restituée sur le spectateur. En effet, les méthodes actuelles réduisent la notion de qualité à un phénomène unidimensionnel ne permettant pas de rendre compte de l'ensemble des effets de la qualité audio et/ou vidéo sur les spectateurs comme l'ont montré Wilson G. M. et Sasse (2000a, 2000b). Le *coût utilisateur*, du point de vue de la fatigue ou de l'effort mental induit par la présence de dégradations, n'est notamment pas reflété par ces méthodes alors même qu'il pourrait agir sur le confort, aspect pouvant être déterminant pour la *qualité d'expérience*. Les mesures psychophysiologiques permettent de recueillir des informations en matière de *coût utilisateur*. Ces mesures dites « objectives » présentent également différents avantages comme celui de ne pas être soumises aux biais inhérents aux évaluations subjectives. Elles offrent également la possibilité de pouvoir recueillir en temps réel l'influence des fluctuations de qualité sur le spectateur sans pour autant interférer avec l'activité de visualisation. Ce dernier aspect permet d'envisager un contexte d'évaluation plus représentatif de conditions réelles d'usage des systèmes de diffusion audiovisuelle notamment en proposant des séquences de test plus longues. En effet, les dix secondes recommandées par les normes UIT pourraient ne pas être suffisantes pour permettre d'observer un phénomène de fatigue ou d'effort mental risquant de conduire, à terme, à un rejet de la technologie ou du système utilisé. Ainsi, l'approche proposée remplit trois objectifs :

- **ajouter des mesures non soumises aux biais des mesures subjectives**,
- **proposer un contexte plus représentatif** des conditions réelles de visualisation en proposant des contenus 2D ou 3D de plusieurs minutes et dont la qualité fluctue,
- **étendre l'évaluation de qualité à l'étude de la *qualité d'expérience*** du spectateur.

Sur la base des travaux de Wilson G. M. et Sasse (2000a, 2000b), l'approche présentée dans ce document consiste donc en une méthode hybride reposant sur l'analyse conjointe de mesures subjectives, physiologiques et oculaires. L'hypothèse générale s'articule de la manière suivante : l'influence de la qualité audiovisuelle sur la *qualité d'expérience* peut être étudiée à la fois sur le plan de la *qualité d'expérience* subjectivement reportée par les spectateurs et sur le plan du *coût utilisateur*, du point de vue de l'effort mental ou de la fatigue, mesuré à partir d'indicateurs toniques de l'activité physiologique ou oculaire du spectateur. L'ajout d'indices de l'expérience subjective et de l'activité physiologique et oculaire doit permettre d'étendre l'évaluation de qualité actuellement utilisée à l'étude plus globale de la *qualité d'expérience*. Cette approche est schématisée par la Figure 5.3 ci-après.

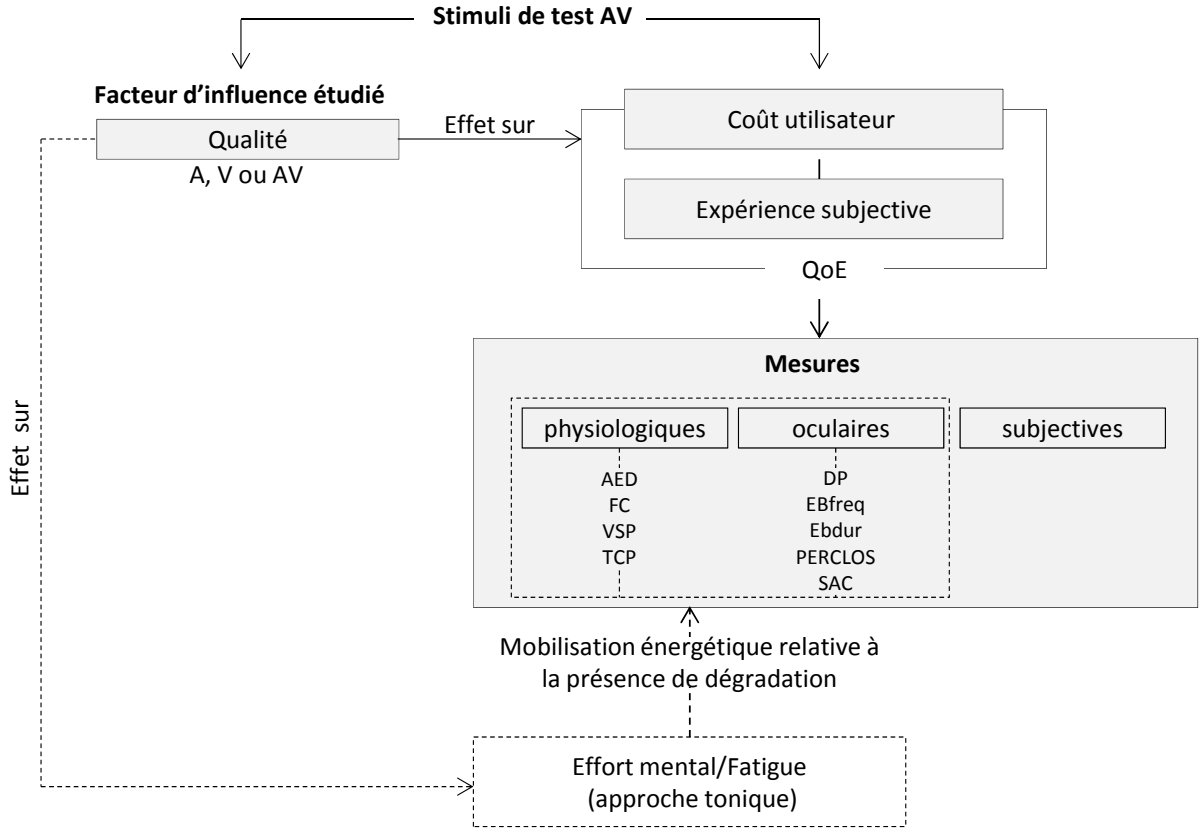


Fig. 5.3. Schéma de l'approche proposée par combinaison des mesures subjectives, physiologiques et oculaires pour l'étude de l'influence de la qualité audiovisuelle sur la *qualité d'expérience* du spectateur.

CHAPITRE VI – EXPERIMENTATION A : ETUDE EXPLORATOIRE

6.1. INTRODUCTION GENERALE

Dans les études de Wilson G. M. et Sasse (2000a, 2000b), le contexte étudié était celui des systèmes de communication multimédia, les séquences de test proposées correspondaient donc à des sons de paroles seuls ou à des sons de paroles accompagnés par la vidéo du locuteur. Au quotidien, le spectateur est confronté à une variété beaucoup plus importante de contextes audiovisuels (AV). Comme indiqué dans les chapitres précédents, la perception de qualité est dépendante du type de contenu (voir sect. 2.2.1, chap. II). Une méthode alternative pour évaluer la qualité perçue et plus généralement la *qualité d'expérience* du spectateur doit considérer différents contextes, plus proches de contextes réels de visualisation. D'ailleurs, la norme UIT-T P.911 (UIT, 1998) propose déjà des contextes variés de séquences de test. L'étude de l'influence de la qualité pour différents types de contenus audiovisuels doit favoriser l'observation des différentes interactions entre contenus et qualité pour en retirer les régularités ou les différences notables (c.-à-d. selon le type de contenu) dont il faudra avoir connaissance. Il est à noter que le terme de visualisation, employé dans la suite du présent document, englobera l'activité de visualisation ainsi que l'activité d'écoute, aucun terme n'existant pour qualifier cette double activité.

Comme indiqué dans la section 1.6.1 (chap. I), la durée des séquences de test proposée par la norme UIT-T P.911 paraît trop courte pour rendre compte des effets éventuels de la qualité liés à une activité prolongée de visualisation. Par exemple, il est tout à fait envisageable qu'une séquence de dix secondes ne produise pas les mêmes effets qu'un contenu de cinq minutes en matière d'effort mental, de fatigue ou même de perception de qualité. Il apparaît donc nécessaire de proposer des séquences suffisamment longues pour étudier ces éventuels effets. Les mesures psychophysiologiques présentent notamment l'avantage de pouvoir observer l'impact éventuel des dégradations au moment où elles se produisent, là où les mesures subjectives seraient majoritairement soumises à des effets mnésiques de primauté et de récence.

Par ailleurs, la norme UIT-T P.911 recommande l'évaluation de la qualité audiovisuelle seule. Diverses études ont cependant montré que les qualités de chacune des modalités s'influencent mutuellement (sect. 2.2.1, chap. II) et impactent la qualité globale perçue par le spectateur (Beerends et de Caluwe, 1999). Ce constat met en évidence l'intérêt de soumettre à l'évaluation non seulement le niveau perçu de qualité audiovisuelle globale mais également les niveaux de qualité audio et vidéo.

6.2. OBJECTIFS

Cette première expérience, de nature exploratoire, propose un corpus de trois contenus de test différents, d'une durée de sept minutes chacun et présentant différents types de dégradations à la fois audio, vidéo et audio-vidéo (dégradation simultanée sur l'audio et la vidéo). L'objectif est d'étudier la manière dont les mesures subjectives classiques d'évaluation de la qualité peuvent être complétées par les mesures physiologiques et oculaires.

6.3. PARTICIPANTS

Trente-trois participants (16 femmes, 17 hommes), entre 22 et 50 ans, avec une audition et une vision normales ou corrigées à la normale (déclaratif), ont participé à l'expérience. Ils étaient rémunérés pour leur participation.

6.4. MATERIEL

6.4.1. CONFIGURATION GENERALE

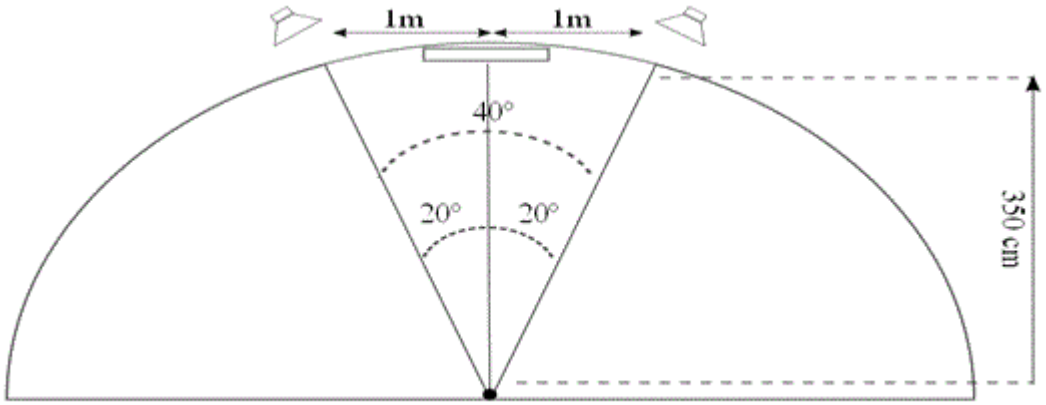


Fig. 6.1. Schéma de la configuration de la salle de test (520×370×285 cm) de l'expérimentation A. La place du participant est représentée par un point noir, l'écran est figuré par un rectangle.

Le test se déroulait dans une pièce insonorisée où les conditions de visualisation et d'écoute recommandées par la norme UIT-T P.911 étaient respectées à l'exception de la chromacité (pour plus de détails voir l'annexe 6-A). Le niveau de luminosité de la salle était identique pour tous les participants (≤ 20 lux -lx-). Les paramètres de brillance et de contraste de l'écran remplissaient les conditions de la norme UIT-R BT.814-2 (UIT, 2007). Un écran LCD Hyundai S465D de 46" (117 cm) full HD (1080p, 16:9) a été utilisé pour afficher les contenus de test en format full HD (1920×1080p). En accord avec la norme UIT-T P.911, la distance de visualisation était fixée à six fois la hauteur de l'écran (350 cm).

Les haut-parleurs (HP) de marque Genelec (modèle 190A) étaient disposés selon une configuration triangulaire, c'est-à-dire qu'ils étaient à la fois équidistants du centre de l'écran (1 m) et équidistants de la tête du participant. La configuration de la salle de test est présentée par la Figure 6.1 ci-dessus, celle-ci respectait les recommandations de la norme UIT-R BS.1286 (UIT, 1997). Le niveau d'écoute respectait le seuil de 80 dBA recommandé par la norme UIT-T P.911 (mesuré à l'aide d'un bruit blanc et d'un sonomètre placé au point d'écoute).

6.4.2. CONFIGURATION TECHNIQUE

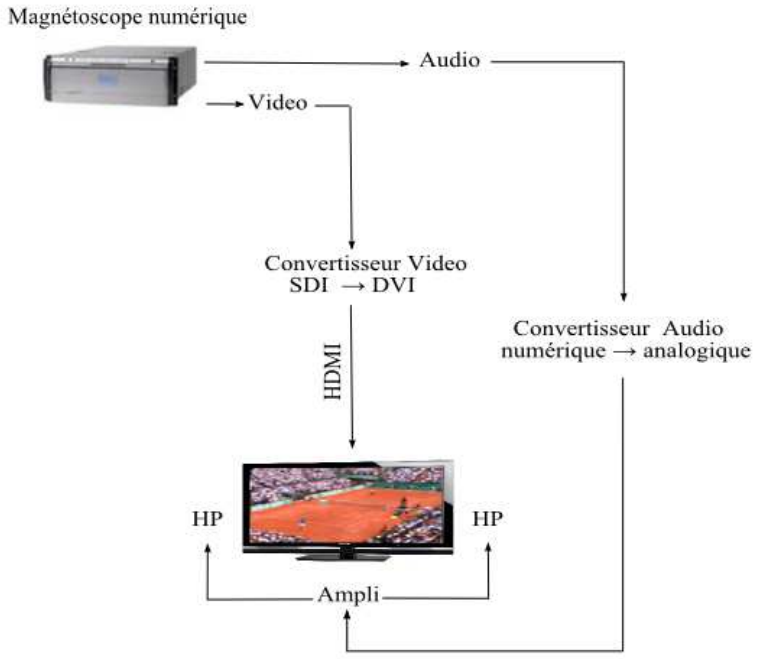


Fig. 6.2. Configuration technique de l'expérimentation A.

La configuration technique de l'expérience A est présentée par la Figure 6.2 ci-dessus. Le stockage et la restitution des contenus AV de test a été effectuée au moyen d'un magnétoscope numérique (Digital Video System -DVS- Pronto2K). Les signaux audio et vidéo étaient ensuite acheminés séparément vers le matériel de restitution vidéo (écran) et audio (HP externes) *via* un amplificateur (SPL 2380). Avant cela, chaque signal était converti au bon format d'entrée, c'est-à-dire d'un signal numérique (signal de sortie du DVS) à un signal analogique (signal d'entrée HP) pour l'audio (convertisseur Apogée Rosetta 800) et d'un signal SDI (signal de sortie du DVS) à DVI (signal de sortie du convertisseur, SDM-875p HD-SDI to DVI) puis HDMI (signal d'entrée écran) pour la vidéo.

L'utilisation du DVS présente l'avantage de pouvoir restituer les signaux audio et vidéo en format full HD (1920×1080) non compressé, c'est-à-dire sans aucune perte de qualité (absence de dégradations liées aux phénomènes de codage/décodage). Cependant, l'utilisation de ce matériel, dans la présente configuration technique, ne permettait pas l'utilisation de différentes listes de lecture des séquences AV de test (fichier appelant les séquences AV selon

un ordre déterminé - playlist -). En conséquence, l'ordre de présentation était fixe et identique pour l'ensemble des participants. Pour obtenir un ordre de présentation randomisé, il aurait été nécessaire de disposer d'autant de trames vidéo (contenant les neuf extraits AV à présenter) que de participants. A titre indicatif, une trame correspondait à environ à 550 Gigabit (Gb), il était impossible, dans le cadre de la configuration mise en place, d'envisager cette solution. Un autre inconvénient résidait dans l'impossibilité de synchroniser le magnétoscope avec les logiciels de mesures physiologiques et oculaires. Par conséquent, le lancement de la vidéo et des différents logiciels de mesures était effectué manuellement. Le magnétoscope ainsi que les écrans de contrôle de la salle étaient installés dans une régie séparée de la salle de test.

6.4.3. RECUEIL DES DONNEES

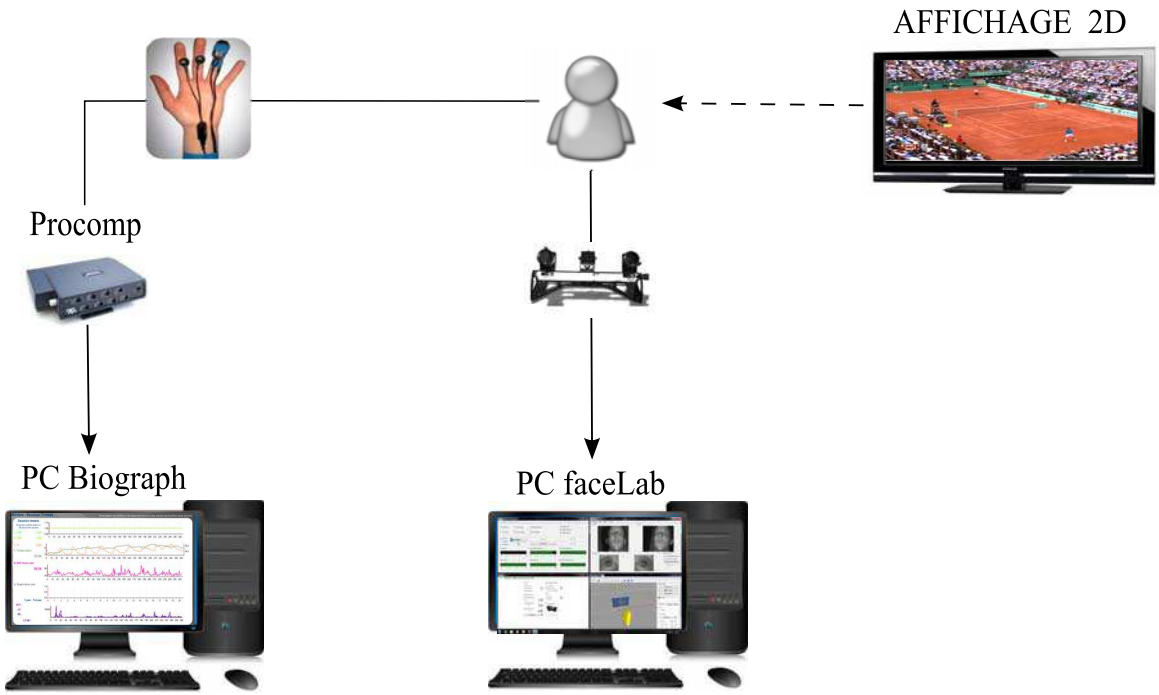


Fig. 6.3. Matériel de recueil des données physiologiques et oculaires.

Comme indiqué dans le chapitre III, les données oculaires étaient capturées par deux optiques (eye tracker faceLAB4™) et acheminées *via firewire* à un ordinateur dédié[12]. Les signaux physiologiques étaient recueillis à l'aide de l'encodeur *Procomp Infiniti* (boitier placé à proximité du participant) et acheminés *via* fibre optique de l'encodeur au logiciel d'enregistrement *Biograph Infiniti* (Thought Technologies™) installé sur un second ordinateur[13]. La Figure 6.3 présente le matériel de recueil des données physiologiques et oculaires. Dans cette expérimentation, le capteur de conductivité électrodermale fourni avec l'outil de mesure *Procomp infiniti* (SC-Flex/Pro, électrodes Ag-AgCl) était utilisé. Celles-ci

[12] Dell Optiplex, Intel Pentium 4 GX620, 3,20 Ghz, ram : 1 Gb, OS: Windows XP (32 bits)

[13] Dell Precision T5400, Intel Xeon E5420, 2,49 GHz, ram : 3,25 Gb, OS : Windows XP (32 bits)

présentent la particularité d'être insérées (cousues) à l'intérieur de sangles velcro permettant leur maintien sur le site d'accueil. L'utilisation de gel pour ces électrodes est déconseillée par le fournisseur. Pour des raisons techniques, les ordinateurs de recueil de mesures étaient placés dans la salle de test. Toutefois, ceux-ci étaient situés à une distance supérieure à 2 m de la place du participant et dissimulés par un rideau, la salle pouvant être scindée en deux espaces distincts. Le bruit inhérent au fonctionnement des ordinateurs a été contrôlé et ne dépassait pas le seuil de niveau de bruit de fond recommandé par la norme UIT-T P.911 (≤ 30 dBA). Pour éviter toute interférence avec les mesures physiologiques, les téléphones portables étaient interdits dans la salle de test.

6.5. STIMULI

Dans cette expérience, trois contenus audiovisuels 2D étaient présentés en format full HD :

- **Opéra *:*** extrait d'une adaptation de *Don Giovanni*,
- **Documentaire** : extrait d'un documentaire sur le boxeur français Jean-Marc Mormeck,
- **Sport** : extrait de la finale de Rolland Garros 2010.

Un aperçu des contenus AV de test est présenté par la Figure 6.4 ci-dessous.

Fig. 6.4. Aperçu des contenus de test de gauche à droite : Documentaire, Opéra et Sport.

Chaque extrait durait sept minutes. Ces trois types de contenus ont été choisis pour couvrir des catégories de programmes présentant des contextes audiovisuels très différents en matière de sémantiques et de choix scénaristiques (dynamique, relation entre les modalités audio et vidéo, changement de plan, *etc*.).

Chaque extrait était présenté trois fois, avec trois conditions de qualité différentes : une condition référence (C0) et deux conditions dégradées (C1 et C2).

La condition C0 correspondait à l'extrait original présenté en qualité source, c'est-à-dire sans compression audio (48Kps, 16 bit) ou vidéo (.avi non compressé).

Une première condition de dégradation (C1) présentait des périodes de désynchronisation (**D**) entre le son et l'image. Chaque période dégradée correspondait à un seuil différent de désynchronisation: - 250 ms, -120 ms, -40 ms, +220 ms et +350 ms, les valeurs négatives représentant un son en avance par rapport à l'image. Ces valeurs ont été choisies pour se situer au-dessus des seuils d'acceptabilité (§ 2.1.3.3, chap. II), à l'exception du seuil fixé à -40 ms devant permettre de tester l'influence d'un niveau de désynchronisation à la limite de la perceptibilité. Chaque période dégradée durait trente secondes et était insérée suivant l'ordre noté ci-dessus. Chacune de ces périodes étaient précédées d'une période de trente secondes

non dégradées (sans désynchronisation). Outre le délai entre l'image et le son, les qualités des signaux audio et vidéo étaient identiques à celle de la condition de référence.

Une deuxième condition de dégradation (C2) présentait des périodes de dégradations du signal A, V ou AV par réduction de débit. Pour rappel, les contenus TVHD sont diffusés avec un minimum de 6 Mbps pour la vidéo et de 128 Kbps pour l'audio. Cinq niveaux de dégradations étaient insérés :

- **A- :** audio compressé (MP3, 16 Kbps)
- **A+ :** audio compressé (MP3, 24 Kbps)
- **V+ :** vidéo compressée (AVC (H264/ MPEG-4 Part 10) - x264, 1500 Kbps)
- **V- :** vidéo compressée (AVC-x264, 1000 Kbps)
- **AV-- :** combinaison des dégradations A- et V-

Comme pour la condition C1, chaque période de dégradation (insérée selon l'ordre donné ci-dessus) durait trente secondes et était précédée par une période de trente secondes non dégradée. La Figure 6.5 présente les patterns des dégradations pour les conditions C1 et C2. Après introduction des dégradations, les segments audio étaient normalisés (homogénéisation du volume) afin d'éviter la présence de sons trop forts ou trop faibles intra ou inter contenu (plage de variation de l'énergie, liée à l'amplitude, comprise dans un intervalle de -1 à 1). Enfin, les contenus étaient décompressés pour pouvoir être diffusé au format full HD.

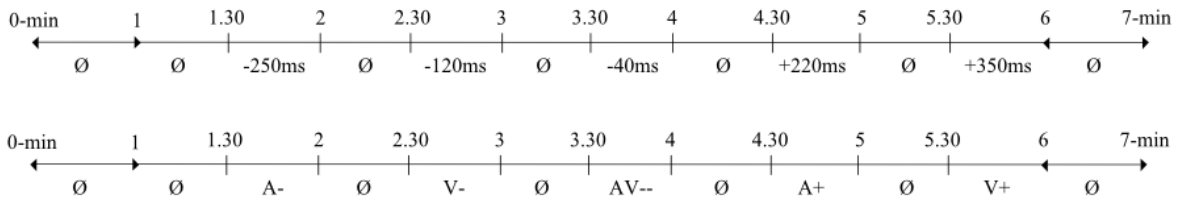


Fig. 6.5. Pattern des dégradations des conditions C1 et C2 où Ø représente une période sans dégradations.

6.6. OBSERVABLES

6.6.1. MESURES SUBJECTIVES

La norme UIT-T P.911 (UIT, 1998) ne proposant pas de question pour évaluer séparément QA et QV, des questions pour l'évaluation séparée de QA et QV ont été extraites de la norme P.920 et ajoutées au protocole d'évaluation. Ainsi, après la visualisation de chaque extrait, les participants devaient évaluer successivement QAV, QV et QA à partir d'une échelle catégorielle en neuf points et cinq items (*Excellent-Bon-Satisfaisant-Médiocre-Mauvais*) comme recommandée par la méthode ACR de la norme P.911 (voir sect. 1.5.1, chap. I). La Figure 6.6 ci-dessous rappelle l'échelle recommandée par la méthode ACR. L'ordre de présentation des questions était conforme à celui proposé par la norme P.920 et était fixe pour l'ensemble des participants (QAV, QV puis QA). L'annexe 6-B présente le questionnaire utilisé.

9 Excellent
8
7 Bon
6
5 Satisfaisant
4
3 Médiocre
2
1 Mauvais

Fig. 6.6. Echelle d'évaluation de la qualité recommandée par la méthode ACR de la norme P.911.

6.6.2. MESURES PHYSIOLOGIQUES ET OCULAIRES

Dans cette expérimentation, les données oculaires recueillies correspondaient à la fréquence et la durée de fermeture de l'œil (EBfreq, EBdur), au PERCLOS et au diamètre pupillaire (DP). Pour les indices physiologiques, la conductance cutanée (AED), la volumétrie sanguine périphérique (VSP), la fréquence cardiaque (FC) et la température cutanée périphérique (TCP) étaient mesurées.

Les fréquences d'échantillonnage choisies pour chaque indice correspondaient aux valeurs maximales par défaut proposées par les outils de mesure, à savoir : 256 Hz pour les données d'AED et de TCP, 2048 Hz pour le VSP et la FC (calculée, *a posteriori*, à partir du VSP) et 60 Hz pour l'ensemble des indices oculaires. Les observables subjectifs et psychophysiologiques pour les différents types de mesures et les outils de recueil utilisés sont synthétisés dans le Tableau 6.1 ci-après.

Tableau 6.1. Synthèse des différents observables et outils de recueil pour chaque type de mesures étudiées.

Type de mesure	Observable	Recueil
SUBJECTIVE	QAV QV QA	Echelles 9 points 5 items (*Excellent-Bon-Satisfaisant-Médiocre-Mauvais)*
PHYSIOLOGIQUE	AED FC TCP VSP	Paire d'électrodes Pléthysmographe Thermistor
OCULAIRE	DP EBdur EBfreq PERCLOS	Eye tracker

6.7. PROTOCOLE

A l'arrivée des participants, les consignes (présentées dans l'annexe 6-C) ainsi que le questionnaire pour l'évaluation de qualité étaient distribués. Directement après cette étape, les capteurs pour l'enregistrement des mesures physiologiques étaient installés sur la main non dominante des participants (paire d'électrodes, pléthysmographe et thermistor). Le placement des différents capteurs a été réalisé comme présenté dans le chapitre III (électrodes pour mesures d'AED placées sur les phalanges médiales de l'index et du majeur). Les mesures enregistrées étant particulièrement sensibles au mouvement, les consignes demandaient donc au participant d'éviter, dans la mesure du possible, tous mouvements du bras ou de la main sur laquelle étaient disposés les capteurs. Il était toutefois possible de repositionner précautionneusement le bras ou la main durant les périodes d'évaluation. L'installation des capteurs peu de temps après l'arrivée des participants avait pour objectif d'une part de stabiliser les mesures physiologiques (différences de température extérieure/intérieure, effort -escalier-, prise de nicotine, *etc.*) et d'autre part, d'habituer le participant au port des capteurs. Après cette étape, le calibrage et la création du modèle de tête individuel étaient réalisés suivant la procédure décrite dans la section 3.6.1 (chap. III). Cette étape durait en moyenne 15 à vingt-cinq minutes. Après le calibrage de l'*eye tracker*, une seconde phase correspondait à l'enregistrement de la baseline durant sept minutes afin de récolter les mesures nécessaires pour établir la condition de comparaison. Durant l'enregistrement de la baseline, les participants avaient pour consigne de se détendre. Aucun stimulus n'était présenté durant cette étape. Après les phases d'installation et d'enregistrement de la baseline, le participant visualisait les trois contenus (Opéra, Documentaire, Sport selon cet ordre, fixe pour l'ensemble des participants), chaque contenu étant présenté successivement trois fois, soit neuf séquences d'une durée de sept minutes chacune. Une pause de deux minutes entre chaque séquence permettait aux participants d'évaluer les niveaux de qualité audio, vidéo et audiovisuelle. La durée totale de passation était d'environ une heure quarante-cinq. La Figure 6.7 présente le déroulement et le chronogramme de l'expérimentation A.

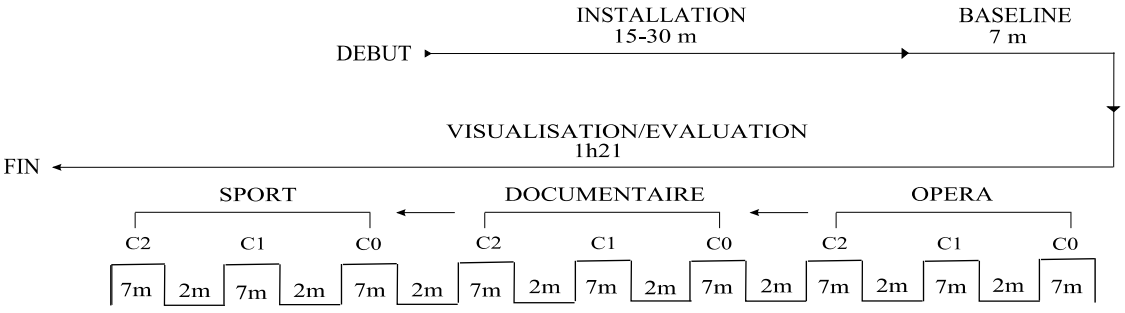


Fig. 6.7. Déroulement et chronogramme de l'expérimentation A. Chaque contenu (7 min), était présenté trois fois, une fois pour chaque condition de qualité (C0, C1 et C2). Après chaque contenu, une pause de 2 min permettait aux participants de compléter le questionnaire.

6.8. HYPOTHESES

Dans cette expérimentation, l'influence des fluctuations de qualité sur l'expérience subjective, physiologique et le comportement oculaire était étudiée. Il était attendu que la présence de dégradations diminue les scores moyens obtenus à l'évaluation des niveaux de qualité audio (MOSA : Mean Opinion Score for Audio), vidéo (MOSV) et audiovisuelle (MOSAV). Deux hypothèses ont été testées pour les mesures subjectives :

- **H0s** : MOSAV, MOSV et MOSA obtenues pour la condition sans dégradations devraient être supérieures à celles obtenues pour les conditions dégradées,
- **H1s** : MOSAV devrait être inférieure à MOSA et MOSV obtenues pour la condition de désynchronisation image/son, la désynchronisation ne dégradant pas les signaux audio et vidéo.

Par ailleurs, il était attendu que les mesures psychophysiologiques réagissent à la présence de l'ensemble ou d'une partie des dégradations introduites. Le traitement du signal dégradé devrait amener le participant à mobiliser plus de ressources (dépenses énergétiques/effort). Une activation du système nerveux autonome et particulièrement de sa branche sympathique (SNS) ainsi que des modifications du comportement oculaire étaient alors attendues. La présence d'un effort répétitif ou prolongé en réponse à la présence de dégradations pourrait également conduire à un état de fatigue. Trois hypothèses ont été testées pour les mesures psychophysiologiques :

- **H0p** : la présence de dégradations audio et/ou vidéo (désynchronisation, réduction du débit) pourrait être à l'origine d'un effort mental supplémentaire (pour décoder et interpréter le message dégradé) puis d'un état de fatigue, visibles à travers des modifications des patterns physiologiques et/ou oculaires,
- **H1p** : un effort mental supplémentaire lié à la présence de dégradations audio et/ou vidéo conduirait à une activation majoritaire du système nerveux sympathique traduite par une augmentation des indices AED, FC, DP et une diminution des indices VSP, TCP, EBfreq et EBdur,
- **H2p** : un état de fatigue consécutif à l'effort mental lié à la présence de dégradations audio et/ou vidéo serait traduit par une augmentation du PERCLOS, de EBfreq et de EBdur.

6.9. RESULTATS

6.9.1. PREPARATION ET REDUCTION DES DONNEES

Pour chacun des participants sélectionnés, environ une heure trente de données ont été recueillies et ce, pour chaque indice. Pour chaque signal mesuré, la première et la dernière minute (pour lesquelles aucune dégradation n'était présente) a été retirée afin de ne

pas tenir compte des « effets de bord » éventuels (surprise, anticipation du début et de la fin du contenu ceux-ci étant d'une durée fixe, réponse d'orientation au début du contenu).

Les mesures physiologiques et oculaires de trois participants n'ont pu être utilisées en raison de problèmes techniques des outils de mesure. En plus de cette première série d'exclusion, d'autres participants ont dû être rejetés en raison du manque de fiabilité de leurs mesures (données manquantes, valeurs aberrantes) pour chaque jeu de données, physiologiques ou oculaires, considérés et traités séparément. Ainsi, un participant supplémentaire a été retiré du jeu de données physiologiques, portant à vingt-neuf le nombre total de participants dont les mesures ont été utilisées. Pour le jeu de données oculaires, six individus ont dû être exclus, portant à vingt-quatre le nombre total de participant retenus. Aucun participant n'a été rejeté pour l'analyse des mesures subjectives. Le nombre de participants dont les mesures ont été retenues pour chaque jeu de données est récapitulé par le Tableau 6.2.

Tableau 6.2. Nombre de participants dont les mesures ont été retenues pour l'analyse statistique à partir des données subjectives, physiologiques ou oculaires.

TYPE DE MESURES	PARTICPANTS RETENUS
SUBJECTIVES	33
PHYSIOLOGIQUES	29
OCULAIRES	24

Les scores subjectifs ont été réduits par la moyenne obtenue pour chaque extrait (Opéra, Documentaire et Sport) et chaque condition (C0, C1 et C2) afin d'obtenir une note MOS pour chaque qualité évaluée, AV (MOSAV), audio (MOSA) et vidéo (MOSV).

Deux autres jeux de données ont été constitués pour les mesures physiologiques et oculaires. Le premier correspondait à la moyenne temporelle de chaque indice, pour chaque participant et pour chacun des neuf extraits soit sur une fenêtre temporelle de cinq minutes (JD-5min). Ce premier jeu de données devait permettre d'observer les éventuelles différences entre les neuf extraits visualisés (effet du contenu, effet global de la condition de qualité). Le second jeu de données a été réalisé de manière identique au premier à la différence que la moyenne temporelle était calculée pour chaque période, dégradée et non dégradée soit une fenêtre temporelle de trente secondes (JD-30s). L'objectif était de pouvoir étudier les éventuelles modifications de l'activité physiologique et/ou oculaire, au sein de chaque extrait et selon la présence, le type et le niveau de dégradation (« V- » par exemple).

Comme indiqué dans la section 3.5 (chap. III), les réponses physiologiques VSP, FC, AED et TCP, sont sujettes à une forte variabilité inter-individuelle. Afin de minimiser cet impact, les moyennes calculées pour les différents indicateurs ont été normalisées par la baseline (*bsl*) pour chaque participant et chaque jeu de données à partir de la formule suivante (Calcanis, Callaghan, Gardner et Walker, 2008 ; Lin *et al.*, 2005) :

$$\text{signal}n = \frac{\text{Moyenne temporelle } signal \text{ (5 min ou 30 s)} - \text{Moyenne temporelle } bsl}{\text{Moyenne temporelle } bsl}$$

Les données normalisées seront notées de la manière suivante : VSP*n*, FC*n*, AED*n* et TCP*n*. La démarche générale pour l'analyse des données physiologiques et oculaires est schématisée par la Figure 6.8 ci-dessous. L'ensemble des figures présentées dans les paragraphes suivants affichera un intervalle de confiance à 95%.

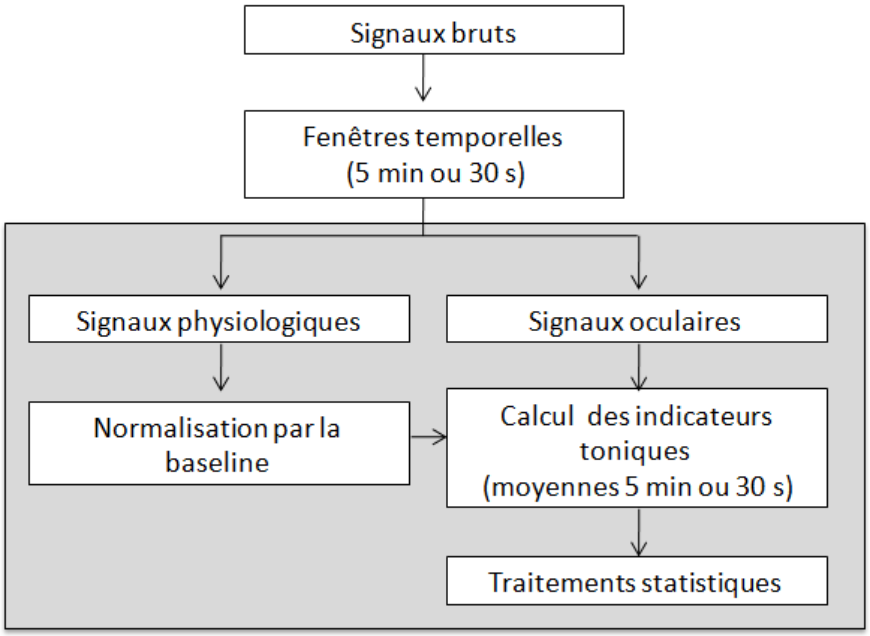


Fig. 6.8. Approche générale pour l'analyse des données physiologiques et oculaires.

6.9.2. MESURES SUBJECTIVES

Selon l'hypothèse H0s, les notes de qualité obtenues pour la condition sans dégradations devaient être supérieures à celles obtenues pour les conditions dégradées. Une analyse multivariée[14] (MANOVA - Multivariate ANalysis Of Variance -) réalisée sur les scores individuels en considérant les variables indépendantes à trois modalités « Contenu » (Opéra, Documentaire et Sport) et « Qualité » (C0, C1 et C2) a confirmé cette hypothèse. Les résultats ont révélé un effet significatif à la fois des facteurs « Contenu » (F (6, 27) = 73,36, p <0,001)

[14] La normalité des données a été vérifiée pour l'expérimentation A ainsi que pour l'ensemble des expérimentations présentées dans la suite du document. Les résultats obtenus (tests Kolmogorov-Smirnov et Lilliefors) montrent qu'une grande partie des distributions étudiées suivaient la loi normale, d'autres non, notamment en raison d'une influence forte du contenu sur la répartition des données. Une approche par analyse de la variance (ANOVA) a donc généralement été privilégiée en raison de sa robustesse aux écarts à la normalité et sa capacité à intégrer, en autres, des facteurs aléatoires.

et « Qualité » (F (6, 27) = 7,43, p <0,001) ainsi qu'une interaction significative entre ces deux variables (F (12, 21) = 17,17, p <0,001). La Figure 6.9 ci-dessous présente les notes MOSAV, MOSV et MOSA obtenues pour chaque contenu et chaque condition de qualité.

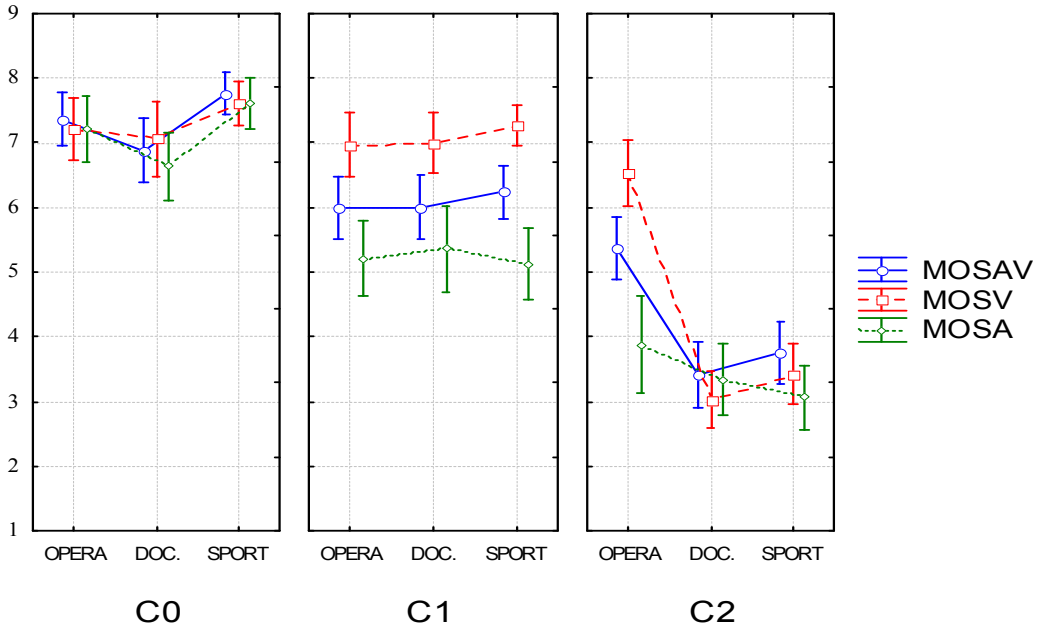


Fig. 6.9. MOSAV, MOSV et MOSA obtenues pour les conditions C0, C1 et C2 pour chacun des contenus visualisés : Opéra, Documentaire (Doc.) et Sport.

La Figure 6.9 indique différents effets. Premièrement, il semble que les notes de QAV correspondent globalement à une moyenne des notes de QA et QV. Ainsi, les participants tiennent compte à la fois de la qualité audio et vidéo perçue pour élaborer leurs notes de jugements de qualité AV. Aucune prédominance de la qualité vidéo ou audio ne semble être mise en avant.

Une seconde hypothèse (H1s) supposait que les notes de qualité AV devaient être inférieures aux notes obtenues pour QA et QV en présence de désynchronisation (C1), la désynchronisation ne dégradant pas les signaux audio et vidéo. Cependant, l'observation des effets de C1 sur la qualité perçue permet de constater une diminution non seulement de QAV mais aussi, et principalement, de QA par rapport à la condition sans dégradations (C0). La qualité audio (présentée sans aucune dégradation) a en effet reçu les scores les plus bas comme si cette modalité était identifiée comme étant porteuse de la dégradation de désynchronisation : le son est en retard ou en avance par rapport à l'image (et non l'image en avance ou en retard par rapport au son).

Par ailleurs, C2 (dégradations A et/ou V) a été perçue comme dégradant plus fortement la qualité que C1, cela est d'autant plus marqué pour les contenus *Documentaire* et *Sport*. L'analyse plus détaillée des effets de C2 sur la qualité perçue a notamment indiqué des scores de QV significativement plus élevés (p < 0,001 d'après un test *post-hoc* HSD de Tukey) pour le contenu *Opéra* que pour les deux autres. Ce constat indique que l'impact de dégradations,

pourtant objectivement équivalentes, sur la perception de qualité est dépendant du type de contenu.

6.9.3. CONCLUSIONS MESURES SUBJECTIVES

Les résultats obtenus pour l'évaluation subjective des niveaux perçus de qualité audio, vidéo et audiovisuelle ont permis de confirmer H0s : les notes obtenues variaient selon les conditions de qualité, les scores pour la condition de référence étaient supérieurs à ceux obtenus pour les conditions dégradées.

En revanche, les résultats n'ont pas permis de confirmer H1s (MOSAV < MOSA et MOSV en présence de désynchronisation), l'audio ayant été considéré comme responsable de la diminution de la qualité audiovisuelle globale. Ce constat exprime une limite des méthodes actuellement recommandées. En effet, en l'absence de question spécifique à la dégradation de désynchronisation, les participants se trouveraient dans l'impossibilité de justifier de manière adéquate leur perception dégradée de QAV. Il semblerait que les échelles actuelles pour l'évaluation de qualité limitent le participant lors de son évaluation de la dégradation par désynchronisation. Cette forme de dégradation est fréquemment rencontrée en contexte réel de visualisation. Une question spécifique portant sur la perception de la désynchronisation pourrait être ajoutée aux normes UIT d'évaluation de la qualité AV afin de favoriser l'étude de son influence tout en évitant de biaiser les réponses apportées aux évaluations des niveaux de qualité perçue A, V ou AV, le participant n'ayant pas la possibilité de juger la désynchronisation autrement qu'à partir des échelles dont il dispose.

Par ailleurs, les participants ont jugé plus sévèrement les dégradations résultant d'une réduction du débit audio et/ou vidéo que les dégradations de désynchronisation. Cette différence est plus importante pour les contenus *Documentaire* et *Sport*, révélant ainsi un effet de contenu. Cet effet a été exprimé par les notes de qualité vidéo, perçue comme bonne pour *Opéra* mais comme médiocre pour les deux autres contenus. Deux explications sont possibles. La première a trait à l'utilisation d'un débit fixe entre les différents contenus pour l'introduction des dégradations vidéo, c'est-à-dire un débit non adaptatif à la complexité du contenu à compresser (texture, niveau de détails, dynamique des mouvements, *etc.*). De ce fait, un débit objectivement équivalent entre contenus peut donner lieu à des dégradations inégales du point de vue perceptif (un seuil identique de débit appliqué à un contenu vidéo « simple » - plus facile à coder - dégraderait moins la qualité perçue qu'un contenu complexe). Or, le contenu *Opéra* présentait une trame vidéo relativement simple par rapport aux autres contenus (peu de détails, peu de mouvements, *etc.*), la diminution du débit entraînant des artefacts moins gênants que pour les autres contenus. Une seconde explication suppose une influence de la modalité dominante. En effet, le contenu *Opéra*, bien que de langue étrangère, proposait essentiellement une action focalisée sur le chant et la musique, peu d'actions étaient présentes du point de vue de l'image (peu de changements de plans, caméra relativement fixe, action frontale localisée sur une scène de théâtre). L'information principale semble donc être véhiculée de manière privilégiée par la modalité audio, détournant

l'attention du participant du signal vidéo et donc de sa qualité. Les contenus *Sport* et *Documentaire* étaient des contenus plus dynamiques, caractérisés par une action principalement axée sur la vidéo (Sport : tennis) voire sur les deux modalités (Documentaire : alternance entre passages d'interviews et scènes d'entraînements au combat). Ce postulat est étayé par les constats de Hands (2004) pour lequel la qualité audio serait dominante pour un contexte AV faiblement dynamique tandis que la modalité vidéo serait dominante pour un contexte AV fortement dynamique, l'attention du spectateur serait alors dirigée vers l'une ou l'autre modalité. Ainsi, l'effet de contenu pourrait être expliqué soit par la notion de modalité dominante, soit par l'impact plus ou moins fort d'une dégradation « objective » en fonction du contenu (détail, dynamique, *etc.*) Le choix des séquences audiovisuelles semble donc déterminant pour l'évaluation de la perception de qualité AV. Une description détaillée des séquences de test audio et vidéo, y compris la relation entre ces deux modalités (modalité dominante par exemple), devrait permettre de mieux comprendre les effets observés.

6.9.4. MESURES PHYSIOLOGIQUES ET OCULAIRES

Chacun des indices physiologiques normalisés (AED*n*, VSP*n*, FC*n* et TCP*n*) et oculaires (DP, PERCLOS, EBfreq et EBdur) a été analysé séparément.

Une première étape a consisté à étudier l'influence du type d'activité sur les mesures physiologiques et oculaires des participants. Pour cela, une série de tests de *Student* a été conduite pour chaque indicateur entre les moyennes obtenues pour la baseline (moyenne unique par participant) et celle du premier contenu visualisé. Une différence significative a été trouvé entre la phase de repos et le premier contenu présenté pour les indicateurs physiologiques AED ($t(28) = 3{,}98$, $p < 0{,}001$) et VSP ($t(28) = -5{,}36$, $p < 0{,}001$). Plus précisément, on observe une augmentation de l'AED et une diminution du VSP lors de la visualisation du premier contenu par rapport à l'état de repos. Des différences ont également été observées pour les indicateurs oculaires (excepté pour EBfreq). Durant la baseline, il était en effet demandé aux participants de regarder devant eux, c'est-à-dire en direction de l'écran (nécessaire pour permettre à l'eye tracker de capter le regard de l'individu et de recueillir les mesures). Cependant, il a été remarqué que les participants avaient du mal à maintenir cette consigne durant les sept minutes que durait la baseline (somnolence, balayage de la pièce, *etc.*) posant la question de la fiabilité des mesure recueillies. De ce fait, les résultats entre repos et activité de visualisation pour les mesures oculaires n'ont pas été pris en compte.

L'hypothèse principale (H0p) supposait que la présence de dégradations audio et/ou vidéo (désynchronisation, réduction du débit) pourrait être à l'origine d'un effort mental (H1p) supplémentaire (pour décoder et interpréter le message dégradé) puis d'un état de fatigue (H2p), visibles à travers des modifications des patterns physiologiques et/ou oculaires. Pour étudier cela, une série d'ANOVAs avec mesures répétées a été réalisée pour chaque indicateur, pour un intervalle de confiance à 95%, sur les deux jeux de données (JD-5min et JD-30s). Pour JD-5min, les ANOVAs conduites tenaient compte des variables indépendantes « Qualité » et « Contenu ». Pour JD-30s, la variable indépendante « Périodes » était ajoutée.

L'ensemble des résultats obtenus est présenté dans l'annexe 6-D. Les analyses *post-hoc* ont été effectuées à l'aide de tests HSD de Tukey. Les principaux résultats sont présentés dans les paragraphes suivants.

La Figure 6.10 ci-dessous présente les moyennes obtenues (à partir des JD-5min) pour chaque contenu et chaque condition de qualité pour les indices DP (fig. 6.10a), VSP*n* (fig. 6.10b) et AED*n* (fig. 6.10c) et qui ont significativement réagi au contenu avec : DP (F (2, 46) = 30,58, p <0,001), VSP*n* (F (2, 56) = 10,95, p <0,001) et AED*n* (F (2, 56) = 7,43, p <0,01).

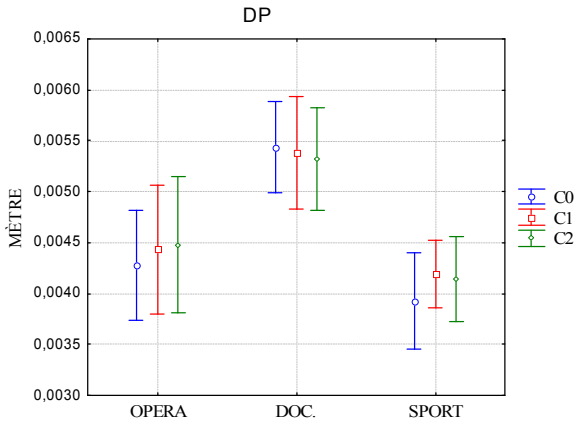


Fig. 6.10a. Moyenne de DP.

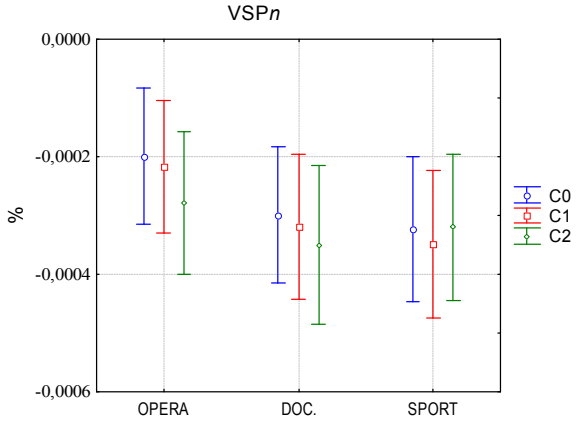


Fig. 6.10b. Moyenne de VSP*n*.

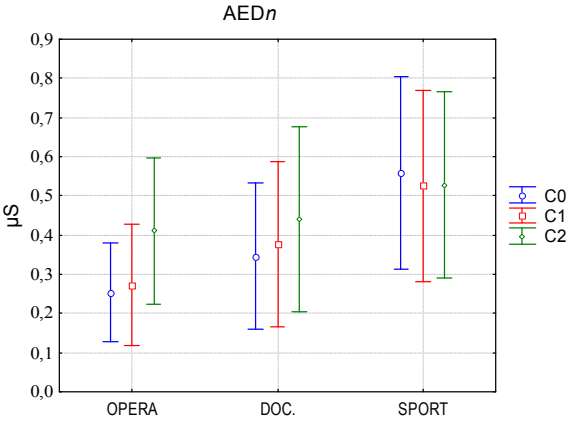


Fig. 6.10c. Moyenne d'AED*n*.

Fig. 6.10. Moyennes obtenues pour chaque contenu Opéra, Documentaire (DOC.) et Sport pour les indices DP (fig. 6.10a), VSP*n* (fig. 6.10b), AED*n* (fig. 6.10c).

Comme l'illustre la Figure 6.10a, le DP moyen obtenu pour le contenu *Documentaire* était significativement plus élevé que pour les deux autres contenus (p <0,001).

Par ailleurs, le VSP*n* moyen était significativement plus élevé lors de la présentation du contenu *Opéra* (p <0,05 entre Opéra et Documentaire et p <0,01 entre Opéra et Sport) malgré une forte variabilité (fig.6.10b).

Les moyennes d'AED*n* (fig. 6.10c) ont également été influencées par le type de contenu : une augmentation significative de l'AED pour le contenu *Sport* par rapport aux deux autres

contenus a été constatée (p <0,05 entre Sport et Opéra et p <0,05 entre Sport et Documentaire).

Les résultats liés à l'AED*n* ont également indiqué un effet de la variable « Qualité » (F (2, 56) = 5,22, p <0,01) et de l'interaction entre les variables « Contenu » et « Qualité » (F (4, 112) = 2,93, p <0,05*)*. En effet, une différence significative entre C2 et les conditions C0 et C1 pour le contenu *Opéra* a été révélé par un test *post-hoc* (avec p < 0,01 entre C2 et C0 et p < 0,01 entre C2 et C1). Une différence significative a aussi été constatée entre C2 et C0 pour le contenu *Documentaire* (p < 0,05).

Les résultats des ANOVAs obtenus pour les périodes JD-5min n'ont pas révélé d'effets significatifs de la variable « Qualité » sur les autres signaux physiologiques ou oculaires.

Les précédents effets, obtenus à l'échelle des contenus (5 min), ont été confirmés à l'échelle des périodes de trente secondes (JD-30s), dégradées ou non, constituant un contenu donné (voir effets principaux des variables indépendantes « Qualité » et « Contenu » donnés en annexe 6-D). Par ailleurs, un effet de la variable « Période » a été constaté pour les indices AED*n* (F (9, 252) = 21,21, p <0,001), FC*n* (F (9, 252) = 5,0, p <0,001), TCP*n* (F (9, 252) = 2,47, p <0,05), DP (F (9, 207) = 4,29, p <0,001), EBdur (F (9, 207) = 6,18, p <0,001) et PERCLOS (F (9, 207) = 7,48, p <0,001).

La Figure 6.11 ci-après, présentant les moyennes d'AED*n* calculées pour chaque période de trente secondes pour chaque contenu (abscisse inférieure) et chaque condition de qualité (paramètre), permet d'observer deux tendances notables de l'AED*n* : premièrement, elle tend à augmenter au fil de l'expérimentation et deuxièmement, elle tend à diminuer durant la visualisation des contenus. La figure indique également une réduction progressive de l'écart des moyennes d'AED*n* entre C0 et C2 au fil du temps (augmentation de l'AED*n* moins forte lors de la présentation du second contenu -Documentaire- et inexistante lors du troisième contenu -Sport-). Ce « tassement » progressif des courbes pourrait indiquer un phénomène d'habituation, l'AED y étant particulièrement sensible.

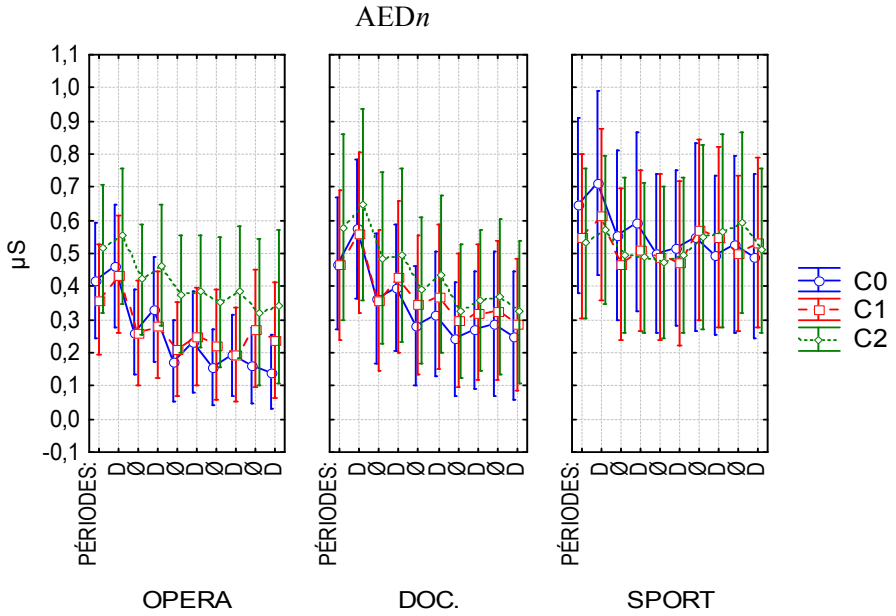


Fig. 6.11. Niveau moyen électrodermal normalisé et obtenu pour chaque période (déradées -D- et non dégradées -Ø-) de chaque contenu (Opéra, Documentaire et Sport) pour les conditions C0, C1 et C2.

Les variations observées sur les autres indicateurs ne semblent pas liées à la présence de dégradations mais plutôt à des réactions physiologiques ponctuelles pouvant, entre autres, résulter de l'influence du contenu ou de la passation du test.

6.9.5. CONCLUSIONS MESURES PHYSIOLOGIQUES ET OCULAIRES

Les résultats obtenus dans cette expérimentation conduisant à plusieurs constats. Tout d'abord, les mesures psychophysiologues sont sensibles **au changement d'activité** (repos *vs.* visualisation). Une augmentation de l'AED*n* moyenne et une diminution du VSP*n* moyen, entre la phase de repos et le premier contenu visualisé, a en effet été observée. Ce résultat semble indiquer une activation physiologique du SNS (mobilisation énergétique) en réponse à l'activité de visualisation. La demande cognitive durant la visualisation serait donc différente de celle sollicitée par l'état de repos. Ce constat indique que les mesures recueillies sont potentiellement capables, dans ce type de contexte, de refléter l'activité cognitive du spectateur.

Les résultats ont également révélé une **influence du contenu** sur les signaux physiologiques et oculaires recueillis. Le niveau de luminosité impacte par exemple le diamètre pupillaire. En effet, la taille moyenne de la pupille était plus importante pour le contenu *Documentaire* (tourné en intérieur) qui était plus sombre que les deux autres contenus visualisés et par conséquent, à l'origine d'une dilatation pupillaire (réponse photo-motrice).

Une **influence de la passation** peut aussi être supposée. Ce postulat s'appuie sur l'observation d'un niveau de VSP*n* moyen significativement plus élevé lors du premier contenu présenté (Opéra). Les différentes conditions de dégradations étaient découvertes au moment de la présentation de ce contenu (pour rappel chaque contenu était présenté trois fois

de suite, une fois pour chaque condition de qualité). Ce qui est intéressant de noter est la diminution de VSP*n* qui était donc observé après la présentation du premier contenu (c.-à-d. durant *Documentaire* et *Sport*). Cette tendance pourrait traduire une activation du SNS liée à un état d'agacement/irritation ou à de l'ennui dû à la présentation répétée de contenus associée à la présentation de dégradations connues. Les participants ont également majoritairement rapportés, durant un temps informel suivant la passation, ne pas avoir apprécié le contenu *Opéra*. La moyenne de VSP*n* durant la visualisation d'*Opéra* pourrait alors refléter un désengagement du participant lorsque ce contenu, peu apprécié, était visualisé pour la seconde et/ou troisième fois.

Par ailleurs, le niveau moyen d'AED*n* a augmenté significativement entre les deux premiers contenus (Opéra et Documentaire) et le dernier (Sport). L'augmentation observée pourrait également exprimer un effet de passation (agacement/irritation, ennui). Calcanis *et al.* (2008) ont supposé qu'une augmentation de l'AED pourrait refléter un phénomène d'ennui voire de désengagement vis-à-vis des séquences audiovisuelles visualisées. Un effet du contenu peut aussi être envisagé pour expliquer l'augmentation de l'AED*n* durant le contenu *Sport* : celui-ci correspondait en effet au contenu le plus dynamique (mouvement de caméra, déplacements rapides et récurrents des joueurs, présence d'éléments textuels : scores, *etc.*). Comme l'ont indiqué Lang A. *et al.* (1999), Lang A. *et al.* (2000), Simons *et al.* (1999) ou encore Yoon *et al.* (1998, sect.4.2.2, chap. IV), l'AED augmente en présence de mouvements ou de changements de plan ou de scène (nombreux pour Sport). Selon ces auteurs, l'apparition de nouvelles informations (changements de plan par exemple) engendrerait une réponse d'orientation liée à une augmentation des ressources attentionnelles pour décoder et interpréter le message. Sur la base de ce postulat, une dynamique élevée (du contenu - personnages- ou de la caméra) ou l'introduction de messages textuels (présence des scores durant le contenu Sport) pourraient donc être à l'origine d'un grand nombre de nouvelles informations à traiter requérant, par conséquent, l'allocation de ressources attentionnelles supplémentaires. Ces ressources nécessiteraient alors des dépenses énergétiques supplémentaires allouées par le système nerveux sympathique et traduites ici par l'augmentation de l'AED. Dans tous les cas, cette augmentation semble traduire une augmentation de l'activation (arousal) soit en réaction à la dynamique soit en réaction à la passation (agacement/ irritation, ennui, désengagement).

En plus de l'effet du changement d'activité, de passation et/ou du contenu, un **effet de la qualité** a été observé. Une augmentation de l'AED*n* moyenne lorsque C2 (c.-à-d. la condition de dégradation par réduction du débit audio et/ou vidéo) était appliquée aux contenus *Opéra* et *Documentaire* (soit le premier et le second contenus présentés au participant) a en effet été constatée. Cette augmentation pourrait traduire une activation du SNS liée à la réalisation d'un effort mental supplémentaire en présence de cette condition de qualité. Le fait que cette influence ne soit pas retrouvée pour le contenu *Sport* pourrait s'expliquer par un effet d'habituation, l'AED étant particulièrement sensible à ce phénomène. Cela tend d'ailleurs à être confirmé par la réduction progressive de l'écart des niveaux d'AED mesurés entre C0 (condition sans dégradation, toujours présentée en premier) et C2 (toujours présentée en dernier) au fil du temps. Ce résultat permet de confirmer l'hypothèse supposant un effort

mental supplémentaire pour le traitement du signal lorsque celui-ci est dégradé (H1p). Cependant, cette hypothèse n'est confirmée par aucun autre indicateur physiologique ou oculaire.

L'absence de différence de l'AED moyenne entre C0 et C1 (c.-à-d. condition de dégradation par désynchronisation) laisse penser que la désynchronisation, bien que perçue et jugée comme dégradant la qualité audio, ne semble pas influencer les mesures physiologiques. Cette absence de réponse pourrait s'expliquer par des effets inégaux des dégradations par désynchronisation et par réduction de débit, les premières dégradant moins la qualité restituée que les secondes comme en témoignent les résultats obtenus à partir des évaluations subjectives. Ainsi, les seuils de dégradations par désynchronisation étaient peut-être trop faibles pour influencer les mesures psychophysiologiques.

Les résultats ont aussi révélé une tendance de l'AED*n* à augmenter au fil du temps mais à diminuer durant la visualisation d'un contenu donné. L'augmentation peut sans doute être attribuée à l'effet de passation (irritation/agacement, *etc*.) discuté ci-dessus. La diminution intra-contenu pourrait en revanche traduire la tendance naturelle de l'AED à diminuer au cours du temps en situation de repos (sect. 3.3.5, chap. III). Entre chaque contenu, les participants devaient évaluer les niveaux de qualité, activité sans doute plus « coûteuse » d'un point de vue cognitif et donc énergétique (effort mnésique, tâche de jugement, lecture/écriture, *etc*.) que la seule tâche de visualisation et se manifestant par une augmentation de l'AED*n* pendant cette phase. Ainsi, l'activité de visualisation, moins coûteuse que la tâche de complétion, refléterait alors un phénomène de repos/détente.

Par ailleurs, malgré le retrait des données enregistrées durant la visualisation de la première minute de chaque contenu, le niveau plus élevé d'AED*n* constaté en début de contenu pourrait tout de même avoir été renforcé par : une activité résiduelle de l'activation liée à la complétion de questionnaire, une adaptation résultant du changement d'activité, c'est-à-dire de la complétion à la visualisation (activation physiologique pour répondre au changement de l'environnement) et à un effet éventuel de surprise/découverte au début de chaque contenu que rien n'annonçait.

De manière générale, l'AED a été l'indicateur le plus sensible aux différentes conditions expérimentales. A l'inverse, les i**ndicateurs du comportement oculaire n'ont pas réagit aux conditions testées** et n'ont donc notamment pas permis de mettre en avant un phénomène de fatigue. Soit les conditions de test n'ont pas engendré de fatigue soit le protocole utilisé n'a pas permis de révéler un tel effet. Pour plus de clarté, les différents résultats et conclusions sont récapitulés par le Tableau 6.3 ci-après.

Tableau 6.3. Récapitulatif des principaux effets des conditions expérimentales sur les mesures psychophysiologiques.

Observations	Effets possibles	Interprétation
↑ AED et ↓VSP entre repos et visualisation	Effet type d'activité	Reflet de l'activité cognitive
↑ AED entre les deux premiers contenus (Opéra/Documentaire) et le dernier (Sport)	Effet passation de test	Augmentation de l'activation (irritation, ennui, désengagement)
	Effet contenu	Augmentation de l'activation (dynamique du contenu Sport)
↓ VSP entre le premier contenu (Opéra) et les deux derniers (Documentaire/Sport)	Effet passation de test	Augmentation de l'activation (irritation, ennui, désengagement)
	Effet contenu	Diminution de l'activation (spécificité du contenu Opéra)
↑ DP pour Documentaire	Effet contenu	Luminosité
↑AED en début de contenu et ↓AED durant la visualisation de chaque contenu	Effet changement d'activité	Effet résiduel, adaptation, surprise, découverte, repos/détente
↑AED pour C2 pour Opéra et Documentaire	Effet qualité	Effort mental

6.10. PISTES D'AMELIORATIONS DU PROTOCOLE

Cette première expérience a montré que la qualité audio et/ou vidéo tend à influencer l'activité physiologique de l'individu notamment à travers la mesure de l'AED. Cependant, ce constat repose uniquement sur les résultats obtenus pour un seul des indicateurs physiologiques et oculaires étudiés. Il se peut donc que cette observation résulte plus de biais contextuels que de la présence d'un effort mental. Le protocole proposé n'a pas non plus permis de mettre en avant un état de fatigue.

L'expérimentation A a souligné la sensibilité des mesures physiologiques et/ou oculaires à des facteurs annexes tels qu'un effet de passation, de changement d'activité ou de contenu qui auraient pu masquer ou atténuer l'observation d'un effet de la qualité. Le protocole proposé nécessite donc d'être réadapté pour diminuer l'impact de ces effets. Les différentes adaptations envisagées sont décrites dans les paragraphes suivants.

6.10.1. SOLUTION FACTEUR PASSATION : VERS UNE SOLUTION DE SYNCHRONISATION

L'évolution constatée de l'AED et du VSP au fil du temps peut s'expliquer par un effet de *passation*. En effet, la présentation successive d'un même contenu (chaque contenu était présenté trois fois de suite, une fois par condition de qualité) a pu conduire le participant à éprouver de l'agacement ou de l'énervement ayant pu masquer certains effets liés aux dégradations de qualité. La présentation des contenus une seule et unique fois semble donc être une des premières améliorations à apporter au protocole. Ce résultat souligne également l'importance de varier l'ordre de présentation des séquences expérimentales. Pour cela, une solution de synchronisation inter-logiciels (voir § 6.10.6 ci-après), c'est-à-dire entre le logiciel

de lecture vidéo, d'enregistrement des mesures physiologiques et celui d'enregistrement des mesures oculaires devra être apportée.

6.10.2. SOLUTION FACTEUR ACTIVITE : CONTENU AMORCE ET AVERTISSEUR

L'adaptation au changement d'activité (repos -baseline- et complétion *vs*. visualisation) a été reflété par les mesures d'AED et de VSP. Cette adaptation peut avoir biaisé les mesures recueillies lors de la présentation du tout premier contenu. L'ajout d'un contenu intermédiaire, entre la baseline et le premier contenu, pourrait permettre d'amortir ce type d'effet et de préparer l'individu à l'activité expérimentale. Ce contenu *amorce* pourrait également faire office de *vanilla* baseline (réalisation d'une tâche cognitive, similaire à la tâche expérimentale, mais requérant un niveau d'effort cognitif moins important) comme défini par Jennings *et al.* (1992, sect. 3.5, chap. III).

Par ailleurs, le fait que le participant n'était pas averti du début de chaque nouveau contenu a également pu avoir une influence sur le pattern physiologique et/ou oculaire des participants (découverte/surprise). Pour pallier ce phénomène, le participant pourrait être informé par un « avertisseur audiovisuel » du début de chaque nouveau contenu (par exemple par un décompte « 5, 4, 3, 2, 1, 0 » associé à des bips sonores).

6.10.3. SOLUTION FACTEUR HABITUATION : VARIATION DES PATTERNS DE DEGRADATIONS

La présentation régulière des dégradations (toutes les 30 s, durant 30 s) au sein de chaque contenu a pu engendrer un phénomène d'habituation. L'impact des dégradations pourrait alors être atténué comme cela tend à être indiqué par la stabilisation progressive des moyennes d'AED au cours de l'expérience (réduction de l'écart des niveaux moyens d'AED entre les conditions 0, 1 et 2 au fil de l'expérience). Un pattern d'introduction différent pour chaque contenu, présenté une seule fois, peut être envisagé.

6.10.4. SOLUTION FACTEUR ENGAGEMENT : MAINTIEN DE L'ENGAGEMENT SUR LA TACHE DE VISUALISATION

La visualisation répétitive des contenus a également pu conduire à un désengagement du participant vis-à-vis de l'activité de visualisation. La présentation unique d'un contenu donné offrirait un contexte plus favorable au maintien du focus attentionnel de l'individu ainsi que de son intérêt sur l'activité de visualisation. De plus, les contenus de test proposés dans l'expérimentation A correspondaient à des extraits, c'est-à-dire à des segments isolés d'une trame narrative plus longue. La présentation du contexte général (description du contexte global) pourrait permettre de mieux situer l'extrait par rapport à son contenu d'origine et d'engager plus largement le participant lors de sa visualisation.

6.10.5. SOLUTION FACTEUR NIVEAU : AUGMENTATION DES DUREES ET SEUILS DE DEGRADATIONS

L'absence majoritaire de réponses significatives physiologiques et oculaires à la présence de dégradations peut également être expliquée par une granularité trop fine du protocole appliqué. En effet, la durée de présentation des dégradations (30 s) ou les seuils utilisés (variables pour une condition de qualité donnée) n'étaient peut-être pas suffisamment élevés pour permettre l'apparition franche de réactions physiologiques et oculaires. Par exemple, les seuils de désynchronisation n'étaient peut-être pas suffisamment élevés pour gêner la fusion multimodale des modalités audio et vidéo (voir chap. II) et par conséquent, générer un effort mental ou un état de fatigue suffisamment important pour être observable à travers les mesures psychophysiologiques recueillies. Cette hypothèse est appuyée par les travaux de Wilson G. M. et Sasse (2000a, 2000b) qui ont montré des réactions physiologiques durant des séquences audio et/ou vidéo présentant des niveaux de dégradations plus élevés (20% de perte de paquets audio, écho, réduction à 5 ips pour la vidéo) et des durées d'application plus longues (2 ou 5 min) de ces dégradations.

6.10.6. SOLUTION FACTEUR MATERIEL : VERS UNE SOLUTION DE SYNCHRONISATION

L'effet supposé du facteur *passation* met en évidence l'importance de varier l'ordre de présentation des séquences expérimentales. Ce constat suppose de devoir s'affranchir de l'utilisation du magnétoscope numérique (DVS). Bien qu'utile par sa capacité à présenter des contenus AV non compressés, il contraint, dans la configuration proposée, à un ordre de présentation fixe entre les participants et à une solution de synchronisation logicielle non optimisée. Une nécessité réside donc dans la mise en place d'une solution de synchronisation entre les logiciels de mesures physiologiques et oculaires et le logiciel de lecture des contenus audiovisuels de test. La synchronisation manuelle a en effet très certainement réduit la fiabilité de la correspondance entre la période du contenu étudiée (5 min ou 30 s) et la fenêtre de données psychophysiologiques devant lui correspondre.

6.10.7. EFFET DE L'ANALYSE : STATISTIQUES COMPLEMENTAIRES

Les analyses statistiques réalisées reposaient uniquement sur l'étude de moyennes. D'autres indicateurs de tendances centrales tels que la médiane ou des indicateurs du domaine fréquentiel, pour l'étude de la variabilité du rythme cardiaque, pourraient être utilisés afin d'apporter une source d'information complémentaire et peut-être plus adaptée à l'étude de l'influence de la qualité sur les mesures recueillies.

6.10.8. SOLUTION FACTEUR CONTENU : POUR UNE CARACTERISATION DES CONTENUS DE TEST

L'influence du contenu a été observée à la fois sur les mesures subjectives et psychophysiologiques. Le contenu est donc un aspect important à prendre en compte pour

interpréter correctement d'une part, les mesures subjectives et d'autre part, les mesures physiologiques et oculaires. Pour cela, une étape de caractérisation des contenus de test semble indispensable. Par exemple, il est possible que la désynchronisation n'ait pas toujours été perçue par les participants. Au-delà des seuils, peut-être insuffisants, de désynchronisation introduits, il est également possible que certaines scènes ne permettent pas ou peu de détecter ce type de dégradation. Cela peut notamment être le cas pour le contenu *Sport*, pour lequel la majorité de l'extrait présentait une information audio qui n'entretenait pas ou peu de lien direct avec l'action présentée par la vidéo (par exemple des commentaires décrivant une action différée). A l'inverse, un décalage entre le son et l'image présent lors d'une scène verbale pourrait dégrader plus largement la qualité perçue. Comme abordé dans le chapitre II, plusieurs études ont montré que la présence de désynchronisation survenant sur des séquences verbales est moins bien tolérée que lorsqu'elle survient sur des séquences non verbales (Hollier et Rimmel, 1998 ; Vatakis et Spence, 2005, voir § 2.1.3.2, chap. II). Les résultats subjectifs ont également suggérer que l'influence d'une dégradation audio ou vidéo sur la qualité perçue est plus ou moins importante selon la modalité dominante du contenu. Les mesures psychophysiologiques semblent aussi avoir été modulées par le contenu notamment par des paramètres de luminosité, de dynamique voire d'intérêt.

Afin d'éviter des interprétations erronées des résultats et de proposer des séquences de test cohérentes et propices à l'expression des effets recherchés, une caractérisation du contenu sur les plans techniques (luminosité, nombre de changement de plans, par exemple.), sémantiques (modalité dominante, compréhension) et hédoniques (plaisir, arousal, intérêt) semble être une étape indispensable pour une bonne maitrise de l'influence du matériel expérimental. Au-delà de cet apport, le recueil de ces informations permettra d'obtenir un retour sur l'expérience subjective du participant avec comme objectif sous-jacent de mieux comprendre le lien entre *qualité d'expérience* et qualité audiovisuelle.

Ces premiers résultats soulignent donc la nécessité d'améliorer le protocole utilisé de façon à proposer un terrain propice à l'expression de l'influence des dégradations notamment à travers des mesures physiologiques et oculaires. Le Tableau 6.4 ci-après récapitule les principales conclusions et adaptations proposées à l'issue de l'expérimentation A.

Tableau 6.4. Récapitulatif des conclusions principales pour chaque type de mesures (Mes.) et des adaptations futures du protocole à l'issue de l'expérimentation A. La désynchronisation est indiquée par la lettre « D ».

Mes.	Conclusions principales	Adaptations proposées
SUBJ.	Effet de la qualité sur les notes MOS	
	Limite du questionnaire (D exprimée au travers de MOSA)	Ajouter une question spécifique à D
	Effet du facteur *niveau* (écart entre C1 et C2)	Augmenter le seuil de D
	Effet du *contenu*	Caractériser les contenus de test
PSYCHOPHYSIOLOGIQUES	Absence d'effets significatifs des différents seuils de dégradations	
	Effet de la Qualité (condition C2 appliquée à Opéra)	
	Effet du facteur *contenu*	Caractériser des contenus de test
	Effet du facteur *passation*	Proposer un ordre aléatoire de présentation/ Présenter une seule et unique fois les contenus
	Effet du facteur *activité*	Ajouter un contenu *amorce* Ajouter avertisseur de début de contenu
	Effet de la qualité masqué ou atténué en raison de :	
	Effet du facteur *habituation* Effet du facteur *engagement* Effet du facteur *niveau* Effet du facteur *matériel* Effet du facteur a*nalyse*	Varier pattern d'introduction des dégradations Favoriser l'engagement sur la tâche Augmenter les durées et les seuils appliqués Proposer solution de synchronisation Varier les approches

L'objectif des expérimentations suivantes est de préparer et de tester un protocole intégrant ces différentes adaptations, l'objectif final étant de pouvoir proposer une méthode alternative valide pour l'étude de l'influence de la qualité restituée des services audiovisuels sur la *qualité d'expérience* du spectateur.

CHAPITRE VII – EXPERIMENTATIONS B : CARACTERISATION ET INFLUENCE DU CONTENU

7.1. INTRODUCTION ET OBJECTIFS

Hands (2004) a souligné l'importance de l'influence du contenu sur l'évaluation subjective de qualité et la nécessité de proposer différents types de contenus de test caractérisés, par exemple, par le niveau de mouvements présent dans la vidéo ou la relation entre les médias audio et vidéo (paroles ou commentaires). Les résultats de l'expérimentation A ont confirmé cette observation en mettant en avant une influence du contenu tant sur les mesures subjectives que psychophysiologiques. L'objectif des expérimentations présentées dans ce chapitre est de proposer un ensemble de descripteurs permettant de caractériser, de la façon la plus complète possible, les contenus de test utilisés. L'impact du contenu sur la perception de qualité et plus globalement sur la *qualité d'expérience* du spectateur pourra donc être étudié à partir des descripteurs spécifiés. La caractérisation doit permettre d'une part, de mieux comprendre la manière dont le contenu, décrit par un certain nombre de critères, influence la perception de qualité et plus largement la *qualité d'expérience* du spectateur et d'autre part, de faciliter l'interprétation des mesures psychophysiologiques.

La norme UIT-T P.911 (UIT, 1998) met à disposition un certain nombre de critères pour décrire les séquences audiovisuelles de test. L'ensemble de ces critères est présenté dans le Tableau 7.1 ci-dessous.

Tableau 7.1. Catégories proposées par la norme UIT-T P.911 pour décrire les contenus audio et vidéo d'une séquence audiovisuelle.

Modalité	Catégorie	Description
Audio	I	Paroles/orateur unique
	II	II Paroles/orateurs multiples
	III	Paroles + musique d'ambiance
	IV	Musique/instrument 1
	V	Musique/instruments multiples
Vidéo	A	Une personne, portrait (tête et épaules) principalement, détail et mouvement limités
	B	Une personne avec des données graphiques et/ou plus de détails
	C	Plusieurs personnes
	D	Données graphique avec pointage
	E	Objet et caméra à contenu cinétique élevé, au-delà de la limite habituellement constatée en télé-vidéoconférence

La description et la classification proposées considèrent l'audio et la vidéo séparément, sans tenir compte du lien entre son et image. De manière générale, la méthode UIT-T P.911 ne tient pas compte des aspects sémantiques (modalité dominante), techniques (changement de plans ou de scènes, mouvement, *etc.*) ou hédoniques (valence et arousal) du contenu audiovisuel. Pourtant, différentes études ont attiré l'attention sur l'influence de ces facteurs comme celle de la dynamique et de la modalité dominante (Hands, 2004), du niveau d'intérêt (Palhais *et al.*, 2012) ou encore de la présence de mouvements et de changements de plans ou de scènes (Lang A. *et al.*, 2000 ; Simons *et al.*, 1999) sur l'évaluation de la qualité perçue et sur les mesures psychophysiologiques. Comme l'a indiqué Hands (2004), une modalité donnée, audio ou vidéo, peut participer de manière plus importante à la note de qualité audiovisuelle en raison de son apport sémantique dominant. La présence de dégradations sur la modalité dominante serait alors d'autant plus gênante.

La perception de qualité repose donc sur différents critères du contenu dont dépendra le jugement du spectateur. Les séquences audiovisuelles doivent être décrites de manière à pouvoir obtenir une interprétation plus précise de la note de qualité attribuée aux signaux audio et/ou vidéo et des influences de la qualité sur la *qualité d'expérience* étudiée à partir de mesures subjectives complémentaires (c.-à-d. autres que la seule note de qualité), physiologiques et oculaires. L'objectif final est de dégager les principaux critères participant à la perception de qualité et plus largement à la *qualité d'expérience* du spectateur.

La caractérisation de contenus de test s'est déroulée selon plusieurs phases. Premièrement, une base exhaustive de descripteurs de contenu a été élaborée avec l'aide d'un expert du domaine de l'audiovisuel (technicien professionnel de l'audiovisuel –société *Digipictoris*, Brest). Dans un second temps, chaque contenu de test a été découpé en unités signifiantes proches d'une analyse plan par plan. Chaque unité de chaque contenu a ensuite été caractérisée à partir du répertoire de descripteurs préalablement élaboré. Une dernière phase a consisté à identifier les descripteurs clés devant constituer le répertoire final.

Le corpus de contenu de test a été enrichi de façon à proposer un plus grand nombre de contextes audiovisuels, à savoir les contenus :

- **Danse :** extrait du ballet *Balé de Rua* (14 min 21),
- **Documentaire** : documentaire entier sur Jean-Marc Mormeck (12 min 25),
- **Opéra** : extrait d'une adaptation de *Don Giovanni* (12 min 36),
- **Sport** *:* extrait de la finale de Rolland Garros 2011 (12 min),
- **Théâtre** : extrait d'une adaptation des *Fourberies de Scapin* (10 min 29).

A la suite de cette caractérisation experte, des séquences de quelques secondes ont été extraites de chaque contenu et présentées à un panel de participants. Leur tâche était de caractériser à leur tour les séquences proposées sur la base, entre autres, de descripteurs utilisés par l'expert (expérimentation B1). Cette étape devait remplir deux objectifs :

- **vérifier la pertinence d'un ensemble de descripteurs** considéré comme plus « subjectifs », afin d'être en mesure d'utiliser la caractérisation experte pour l'intégralité des contenus,
- **étudier la pertinence de descripteurs supplémentaires** davantage liés à la *qualité d'expérience* spectateur (plaisir ou intérêt par exemple).

Enfin, les interactions entre contenus et qualité perçue ont été étudiées à la lumière de ces descripteurs (expérimentation B2).

7.2. EXPERIMENTATION B1 : CARACTERISATION DES CONTENUS

7.2.1. SELECTION DES DESCRIPTEURS

Un contenu peut être décrit à partir de différentes catégories de descripteurs, par exemple, des descripteurs techniques relatifs au choix de réalisation tels que le nombre de changements de plans, la dynamique de caméra (zooms, travellings, *etc.*) ou des descripteurs sémantiques tels que la modalité dominante, le niveau de compréhension ou encore la quantité d'information perçue. La caractérisation experte a été réalisée à l'aide de vingt-huit descripteurs pouvant être regroupés au sein de deux grandes catégories : les descripteurs techniques et les descripteurs sémantiques. Un exemple du support utilisé par l'expert pour décrire une séquence donnée est apporté par l'annexe 7-A.

Certaines nomenclatures existent pour décrire des contenus audiovisuels. Notamment la norme MPEG7 (ISO/IEC, 2004) propose une description standard de contenus multimédias dans le cadre d'applications de recherches étendues de documents archivés. Elle fournit notamment un ensemble de descripteurs dit de bas-niveau d'abstraction tels que le mouvement de caméra (fixe, panoramique -rotation horizontale-, travelling -mouvement transversal horizontal-, zoom, *etc.*), la texture (niveau de détail), la température de couleur ou encore la dynamique, définie comme la notion intuitive de l'intensité ou du rythme de l'action d'une séquence vidéo. Des descripteurs de plan ont également été proposés par Amiar (1995) : durée, angle de vue (plongée, contre-plongée), mouvements de caméra, cadrage (gros plan, plan d'ensemble, *etc.*), profondeur de champ (flou, courte, grande, *etc*.).

Le choix des **descripteurs techniques** s'est appuyé sur l'ensemble de ces spécifications. Au total, treize descripteurs techniques ont été retenus : niveau de détail (faible-modéré-fort), température de couleur (chaude, jour, froide), luminosité (faible-modérée-forte) et caractéristiques de caméra (générale, mobilité, angle de prise de vue -horizontal et vertical-, cadrage, nombre de *cuts*, zoom, rotation caméra, profondeur de champ, angle de vue, pour plus de détails voir annexe 7-A).

La sélection des **descripteurs sémantiques** a été réalisée sur la base des descripteurs proposés par la norme MPEG7 ainsi que ceux suggérés par Amiar (1995). Cet auteur propose notamment des paramètres scénaristiques (intérieur/extérieur, jour/nuit, visuel/dialogue,

action-tension/inaction-immobilité, nombre de personnages, intime/collectif/public), des caractéristiques audio (parole, bruit, musique) ou encore la qualification des relations image/son (son diégétique : son *in* ou hors-champ ; son extra-diégétique : son *off*, ces relations seront mieux définies ci-après). Par ailleurs, selon Zettl (1991, cité par Simons *et al.*), le mouvement (*motion*) d'un film ou d'un contenu télévisuel peut être décrit à la fois comme le mouvement d'un objet présent à l'image (balle de tennis par exemple), le mouvement des caméras (travelling, zoom, panoramique, inclinaison, *etc.*) et le mouvement de la séquence (changement de plans par utilisation de *cut* ou tout autre moyen de transition). Au total, quinze descripteurs sémantiques ont été retenus : modalité dominante (audio, vidéo, audiovisuelle : sur la base des résultats de l'expérimentation A et des constats de Hands, 2004), présence de mouvements, présence d'informations textuelles, dynamique de contenu (faible-modérée-forte), dynamique caméra (faible-modérée-forte), expression sonore (parole, musique, bruit), type de parole (dialogue-monologue, commentaires, chant), relations image/son (son *in*, *off* ou hors-champ) ainsi que l'ensemble des critères scénaristiques proposés par Amiar : intérieur/extérieur, jour/nuit, clair-sombre, visuel/dialogue, intime/collectif/public, nombre de personnages, action/inaction.

Les vingt-huit descripteurs sémantiques et techniques ont été utilisés par l'expert pour caractériser la totalité des contenus du corpus. Pour permettre ce processus, chaque contenu a été segmenté par l'expert en différentes séquences de temps (proche d'une analyse plan par plan). Chacune de ces séquences peut être considérée, comme défini par Goliot-Lété et Vanoye (1993, p. 28, cité par Amiar), comme une unité de sens, c'est-à-dire une suite de scènes qui ne se déroulent pas forcément dans le même décor, mais qui forme un tout avec un sens lui étant propre.

La caractérisation experte a permis de faire émerger neuf descripteurs sémantiques et techniques principaux. L'ensemble des descripteurs n'a en effet pas été retenu en raison de la redondance de certaines informations ou de la granularité parfois trop fine de certains descripteurs (par exemple le type de cadrage, voir annexe 7-A). Les descripteurs techniques qui ont été sélectionnés sont :

- **la luminosité** (faible, modérée, forte)**,**
- **la température de couleur** (chaude -orangée-, jour -lumière blanche-, froide -bleutée- sect. 1.4, chap. I),
- **la dynamique caméra** (faible, modérée, forte : regroupe les différents mouvements de caméra - travellings, rotations, zooms, *etc.* – incluant *cuts*/changements de plan),
- **le niveau de détail** (faible, modéré, fort).

Les descripteurs sémantiques qui ont été sélectionnés sont:

- **le rapport audiovisuel** ou **diégèse** (son *in*, *off* ou hors-champ),

Dans ce document, la notion de *diégèse*[15] fera référence à l'ensemble des sons pouvant être qualifié de son *in*, *off* ou hors-champ. Deux types de sons *in* (sons diégétiques c.-à-d. se déroulant dans le même espace-temps que l'action) peuvent être distingués : dans le champ ou son *in* (accompagne l'action et entendu par les personnages de la scène[16]) et hors-champ (hors de la scène -hors du champ de la caméra et donc du spectateur- mais entendu par les personnages[17]). Un son *off* (son extra-diégétique) se définit par un son en-dehors de l'espace-temps de l'action et qui n'est pas entendu par les personnages de la scène mais par le spectateur (voix *off* de narration[18] ou musique *off*[19]). Dans les études suivantes, tous sons *in* (dans le champ et hors-champ) seront considérés comme diégétiques tandis que les sons *off* seront considérés comme extra-diégétiques.

- **l'expression sonore** (parole, musique, bruit),
- **le nombre de personnages** (faible ≤ 2, modéré 2 à 5, fort ≥ 5),
- **la dynamique de contenu** (faible, modérée, forte)
- Le terme de contenu est accolé à la notion de dynamique dans l'intention d'établir une distinction nette avec le premier descripteur de dynamique relatif aux mouvements de caméra (descripteur technique). La *dynamique du contenu* réfère à l'action des personnages ou des objets.
- **la modalité dominante** (A, AV, V).
  La *modalité dominante* peut se définir comme la modalité porteuse de l'information primordiale et sans laquelle la compréhension de la séquence serait mise à mal.

Expérimentation B1 : Contenu et expérience spectateur

7.2.2. OBJECTIFS

Afin de pouvoir considérer la caractérisation réalisée par l'expert comme pertinente, des séquences ont été extraites de chacun des contenus du corpus pour être soumises à l'évaluation d'un public « naïf ». L'objectif ici est de pouvoir observer une concordance entre les annotations de l'expert et celles des naïfs à partir d'un échantillon de séquences (les extraits entiers n'ont pas été présentés en raison du coût extrêmement élevé, en temps et en

[15] La notion de diégèse a été créée et définie par Souriau (1951[15]) comme «tout ce qui est censé se passer, selon la fiction que présente le film ; tout ce que cette fiction impliquerait si on la supposait vraie».

[16] Par exemple, paroles prononcées par un personnage (interview de Jean-Marc Mormeck) ou bruits issus des actions d'un personnage (sons associés à une scène de combat).

[17] Bruits de pas d'un personnage non visible ni pour les personnages de la scène en cours ni pour le spectateur.

[18] Typiquement, voix d'un locuteur qui commente la scène sans que ce dernier n'existe pour les personnages : commentaires sportifs.

[19] Orchestre d'opéra

concentration, nécessaire à la réalisation de la caractérisation -analyse par unités de signification c.-à-d. pratiquement plan par plan-) afin de pouvoir utiliser la caractérisation experte réalisée sur l'ensemble du corpus.

Cependant, la totalité des descripteurs issus de la caractérisation experte n'a pas été soumise à l'annotation naïve en raison de la nature immuable de certains descripteurs. Cela concernait les descripteurs *Détails*, *Mouvement caméra*, *Relations AV, Expression sonore* et *Nombre de personnages*. Par exemple, le nombre de personnage présent dans une scène ne variera pas selon les perceptions individuelles. Ainsi, seuls quatre descripteurs : *Modalité dominante*, *Couleur*, *Luminosité* et *Dynamique de contenu* ont été évalués à la fois par l'expert et par les participants. Ces derniers sont considérés comme potentiellement variables selon l'individu qui les perçoit.

Un contenu pourrait également être décrit par sa qualité hédonique, c'est-à-dire son niveau d'intérêt ou sa valence par exemple. Dans l'intention de couvrir ces notions propres à l'expérience du spectateur, cinq descripteurs ont été ajoutés. Afin de distinguer ces descripteurs des précédents, ils sont qualifiés de haut-niveau d'abstraction tandis que les descripteurs utilisés pour la caractérisation experte sont qualifiés de bas-niveau.

Les descripteurs de haut-niveau incluent trois descripteurs relatifs à la qualité *hédonique* d'un contenu à savoir l'*Intérêt*, le *Plaisir* et l'*Arousal*, et deux descripteurs qualifiés de sémantiques : *Compréhension* et *Quantité d'information perçue*. Les descripteurs *Intérêt*, *Compréhension* et *Quantité d'information* ont été évalués à partir des niveaux : *faible*, *modéré* ou *fort*. Les descripteurs de plaisir et d'arousal, reconnus pour être les dimensions décrivant le mieux une émotion (Lang. P. *et al.* 1993), étaient annotés au moyen des échelles picturales SAM (Self-Assessment Manikin, voir sect. 4.1, chap. IV).

L'ensemble des descripteurs, de bas et haut-niveau, doit permettre à terme de mieux comprendre les interactions possibles entre contenu et perception de qualité ainsi qu'entre le contenu et la *qualité d'expérience* globale du spectateur. L'ensemble des descripteurs annotés par l'expert et/ou par les participants est récapitulé dans Tableau 7.2 ci-dessous.

Tableau 7.2. Récapitulatif des différents descripteurs utilisés par l'expert et/ou par les participants naïfs, ainsi que leurs échelles d'annotations, classés selon les catégories Technique, Sémantique ou Hédonique et selon leurs niveaux d'abstraction.

Descripteur	Echelle	Catégorie	Niveau	Annotation Experte	Annotation Naïve
Dynamique caméra	faible-modérée-forte	Technique	Bas	X	
Détail	faible-modéré-fort	Technique	Bas	X	
Nb personnages	faible-modéré-fort	Sémantique	Bas	X	
Relation AV	in/off/hors-champ	Sémantique	Bas	X	
Expression sonore	parole-musique-bruit	Sémantique	Bas	X	
Luminosité	faible-modérée-forte	Technique	Bas	X	X
Couleur	chaude-jour-froide	Technique	Bas	X	X
Modalité	A, V, AV	Sémantique	Bas	X	X
Dynamique contenu	faible-modérée-forte	Sémantique	Bas	X	X
Compréhension	faible-modérée-forte	Sémantique	Haut		X
Quantité d'information	faible-modérée-forte	Sémantique	Haut		X
Intérêt	faible-modéré-fort	Hédonique	Haut		X
Valence	9 niveaux (SAM)	Hédonique	Haut		X
Arousal	9 niveaux (SAM)	Hédonique	Haut		X

7.2.3. PARTICIPANTS

Vingt huit participants *naïfs* (9 femmes, 19 hommes) entre 23 à 50 ans ont participé à cette expérimentation.

7.2.4. MATERIEL

7.2.4.1. CONFIGURATION GENERALE

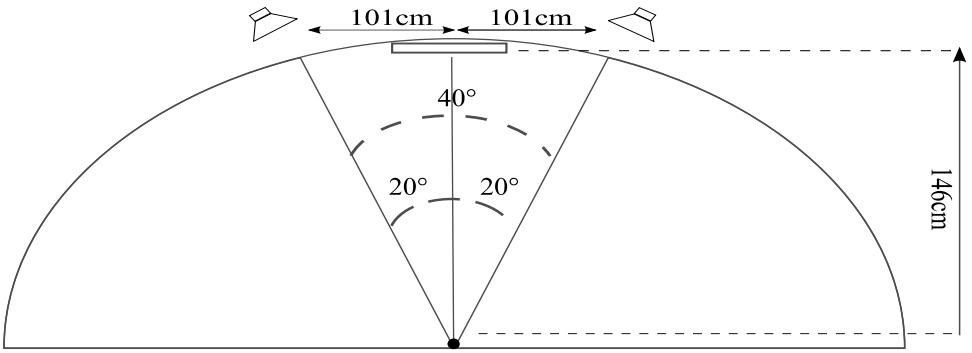


Fig. 7.1. Schéma de la configuration de la salle de test (193×376×505 cm) de l'expérimentation B1. La place du participant est représentée par un point noir, l'écran est figuré par un rectangle.

Les conditions de passation (salle, luminosité, *etc*.) étaient identiques à celles de l'expérimentation A (annexe 6-A). Concernant l'affichage, un écran LCD de 24 " (61 cm), full HD (1080p, 16:9) de type Acer modèle GD245HQ a été utilisé. La distance de visualisation, en accord avec la norme UIT-T P.911, a été fixée à 146 cm soit cinq (4,95) fois la hauteur de l'écran. L'ensemble des séquences de test utilisées a été présenté en format *.avi* non compressé (full HD, 1080p).

Des haut-parleurs Genelec de modèle 8040A ont été réglés à une hauteur de 94,5 cm et placés de manière équidistante du centre de l'écran (101 cm) et de la tête du participant (225 cm). La Figure 7.1 ci-dessus présente la configuration de la salle de test établie en conformité avec les recommandations de la norme UIT-R BS.1286 (UIT, 1997). De manière identique à l'expérimentation A, le volume sonore, mesuré à la tête du participant pour simuler les conditions réelles d'écoute, a été paramétré de manière à se situer autour de 80 dBA comme recommandé dans la norme UIT-T P.911.

7.2.4.2. CONFIGURATION TECHNIQUE

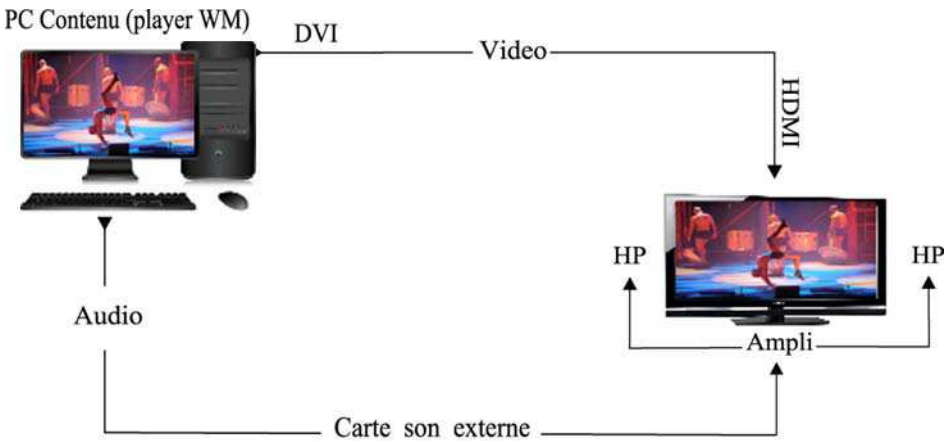


Fig. 7.2. Configuration technique de l'expérimentation B1.

Les séquences AV étaient stockées sur un ordinateur[20] suffisamment puissant pour restituer des contenus full HD non compressés. Comme indiqué par la Figure 7.2, le signal vidéo était acheminé de l'ordinateur vers l'écran *via* une connectique DVI-HDMI du système de diffusion (PC Contenu : sortie DVI) au terminal télévisuel (entrée HDMI). La restitution du signal audio sur les HP était effectuée au moyen d'une carte son externe (Terratec Auréon 5.1 MKII) et un amplificateur (SPL 2380). Les séquences audiovisuelles ont été diffusées *via* le *player* multimédia Windows Media (WM). Cette configuration permettait l'utilisation de *playlist*, par conséquent, les séquences ont pu être présentées avec un ordre aléatoire, différent pour chaque participant.

[20] Dell Precision T5500, intel Xeon x5687, 3,60 Ghz, ram : 6 Gb, DDR3, Hard Drive (x2) 600 Gb SAS (15000RPM), OS : windows 7 (64 bit), Carte graphique NVIDIA Quadro 600, 1 Gb

Les questions s'affichaient et les participants répondaient sur une tablette tactile (SESOL Co., Ltd.), permettant l'enregistrement automatique des réponses sur un ordinateur annexe. L'ensemble de l'installation, c'est-à-dire l'ordinateur utilisé pour la lecture des séquences et celui utilisé pour enregistrer les données était placé en régie.

7.2.5. STIMULI

Dans cette expérience, vingt séquences de huit à dix secondes présentées en format 2D full HD 1080p (format .avi non compressé et audio 16 bit, 48 Kps) constituaient le corpus de test.

Deux paires de séquences ont été extraites de chaque contenu, une paire étant caractérisée par un descripteur sémantique particulier avec chaque séquence représentant un niveau particulier du descripteur sémantique. La distribution des paires selon le descripteur à représenter est visible dans l'annexe 7-B. Par exemple, la séquence 1 de la paire A du contenu *Documentaire* représentait le mode *musique* du descripteur *Expression sonore* tandis que la séquence 2 représentait le mode *Parole*. Afin de couvrir l'ensemble des modes de chaque descripteur, une autre paire de séquences représentait le descripteur *Expression sonore*. Ainsi, la séquence 1 de la paires B du contenu *Théâtre* représentait le mode *Parole* tandis que la séquence 2 représentait le mode *Bruit*. Au total, un descripteur était donc représenté par deux paires (quatre séquences), chaque paire étant issue d'un contenu différent. Dans la mesure du possible, les modes des descripteurs sémantiques non représentés par la paire devaient être identiques pour chaque séquence afin de ne faire varier que le descripteur exprimé. Les caractéristiques techniques étaient toujours identiques entre les séquences d'une paire donnée. Comme réalisé dans l'expérimentation A, le volume sonore entre les séquences de test a été homogénéisé pour éviter la présence de disparités importantes entre les différentes séquences AV.

7.2.6. OBSERVABLES

En plus de l'évaluation des neufs descripteurs retenus pour cette phase (quatre de bas-niveau : modalité dominante, dynamique de contenu, luminosité, température de couleur et cinq de haut-niveau : intérêt, plaisir, arousal, compréhension et quantité d'information), le questionnaire présentait également trois échelles dédiées à l'évaluation des qualités audiovisuelle, vidéo et audio. Les échelles étaient identiques à celles utilisées lors de l'expérimentation A. Cela portait le nombre total d'observables à douze.

7.2.7. PROTOCOLE

Les participants visualisaient un total de vingt séquences audiovisuelles. Entre chaque visualisation, une période de deux minutes permettait l'évaluation des séquences sur la base des douze descripteurs proposés. Le questionnaire utilisé est présenté dans l'annexe 7-C. Au total, la passation du test durait environ quarante-cinq minutes.

7.2.8. HYPOTHESES

La présente expérimentation devait permettre de vérifier la pertinence d'un certain nombre de descripteurs : soit commun avec ceux utilisés par l'expert, soit en lien avec des influences plus individuelles. Deux types de résultats étaient attendus : une cohérence entre l'annotation experte et naïve (pour les descripteurs en commun) ; un effet de la séquence sur l'évaluation de l'ensemble des descripteurs ainsi que sur l'évaluation de qualité. Ces hypothèses peuvent être résumées de la manière suivante :

H0 : observation d'une cohérence entre l'annotation experte et naïve
H1 : effet de la séquence sur les descripteurs des catégories Hédonique, Sémantique, Technique ainsi que sur les notes de Qualité

7.2.9. RESULTATS

Les figures présentées ci-après présenteront un intervalle de confiance à 95%.

7.2.9.1. ANNOTATION EXPERTE *VS*. NAÏVE

Pour permettre la comparaison entre la caractérisation de l'expert et celle des naïfs, les séquences évaluées par les participants ont été recodées selon le mode obtenu pour chaque descripteur (modalité la plus fréquente, voir annexe 7-D). Un tableau de contingence a ainsi pu être réalisé pour chaque descripteur évalué en commun avec l'expert : *Modalité dominante*, *Dynamique de contenu*, *Température de couleur* et *Luminosité*. Les tableaux de contingence obtenus sont présentés dans le Tableau 7.3 ci-après.

Tableau 7.3. Tableaux de contingence obtenus pour les annotations expertes et naïves (selon le mode) réalisées pour chacune des vingt séquences à partir des descripteurs Modalité, Dynamique, Couleur et Luminosité. Les colonnes correspondent à l'annotation experte tandis que les lignes correspondant au mode issu des réponses du panel de naïfs. Les effectifs ainsi que leurs traductions en pourcentage sont indiqués.

Modalité

Modes	AV	V	A	Total
AV	0	2	0	2
	0%	*100%*	*0%*	*100%*
V	2	7	1	10
	20%	*70%*	*10%*	*100%*
A	0	0	8	8
	0%	*0%*	*100%*	100%
Total	2	9	9	20
	10%	*45%*	*45%*	*100%*

Dynamique

Modes	Faible	Modérée	Forte	Total
Faible	6	1	0	7
	85,71%	*14,29%*	*0%*	*100%*
Modérée	5	4	0	9
	55,56%	*44,44%*	*0%*	*100%*
Forte	0	0	4	4
	0%	*0%*	*100%*	*100%*
Total	11	5	4	20
	55%	*25%*	*20%*	*100%*

Couleur

Modes	Chaude	Jour	Froide	Total
Chaude	4	3	4	11
	36,36%	*27,27%*	*36,36%*	*100%*
Jour	0	5	0	5
	0%	*100%*	*0%*	*100%*
Froide	0	4	0	4
	0%	*100%*	*0%*	*100%*
Total	4	12	4	20
	20%	*60%*	*20%*	*100%*

Luminosité

Modes	Faible	Modérée	Forte	Total
Faible	6	0	0	6
	100%	*0%*	*0%*	*100%*
Modérée	6	1	1	8
	75%	*12,50%*	*12,50%*	*100%*
Forte	0	3	3	6
	0%	*50%*	*50%*	*100%*
Total	12	4	4	20
	60%	*20%*	*20%*	*100%*

Une première hypothèse supposait l'observation d'une cohérence entre les annotations experte et naïve. Afin de tester cette concordance, un test de Kappa de Cohen a été réalisé. Les résultats sont présentés par le Tableau 7.4 ci-dessous. Ils ont indiqué un accord entre expert et naïfs pour les descripteurs *Modalité* et *Dynamique*. Comme l'indique les tableaux de contingences ci-dessus, les participants ont majoritairement répondu de manière identique à l'expert pour l'annotation des niveaux de luminosité « fort » ou « modéré », en revanche, les annotations concordent moins pour le niveau « faible ». Les séquences annotées par un niveau faible de luminosité par l'expert ont été caractérisées par un niveau faible ou modéré par les participants. Il semble que ce soit l'annotation du descripteur de température de couleur qui ait vraiment posé problème. En effet, une faible voire une absence de concordance peut être observée entre les annotations de l'expert et celles des naïfs pour les modalités « Jour » et « Froide ». Il semblerait que ces termes, réservés au monde de l'audiovisuel, prêtent à confusion et soit peu ou mal compris par un public non expert.

Tableau 7.4. Résultats des tests de Kappa réalisés entre l'annotation experte et naïve pour chacun des descripteurs évalués.

Naïfs *vs.* Expert	**Test**	**Valeur**	**Signification**
Modalité	K	0,57	*< 0,001*
Dynamique	K	0,54	*< 0,001*
Luminosité	K	0,27	*0,059*
Couleur	K	0,21	*0,098*

7.2.9.2. CARACTERISATION NAÏVE DES SEQUENCES

Les spectateurs ont caractérisé, à partir des neufs descripteurs (de haut et bas-niveau) hédoniques, sémantiques et techniques, les vingt séquences visualisées. L'hypothèse H1 supposait un effet de la séquence sur les différents descripteurs annotés. Une ANOVA considérant la variable indépendante « Séquence » et le facteur aléatoire « Participant » a donc été réalisée pour chacun des descripteurs évalués à l'exception des variables nominales « Couleur » et « Modalité » pour lesquelles un Khi-deux de Pearson a été conduit. L'ensemble des résultats obtenus est présenté dans l'annexe 7-E. Un effet systématique de la variable

« Participant » a été constaté avec $p < 0{,}001$. Les résultats ont par ailleurs permis de confirmer H1, en effet, une influence significative de la séquence a été observée pour la totalité des descripteurs. Il semble donc que les descripteurs proposés aient été suffisamment explicites pour les participants (pas de difficulté de compréhension).

DESCRIPTEURS DE LA CATEGORIE TECHNIQUE

Les Figures 7.3 et 7.4 ci-dessous présentent respectivement les résultats obtenus pour les descripteurs *Luminosité* (moyennes) et *Couleur* (répartition par effectif).

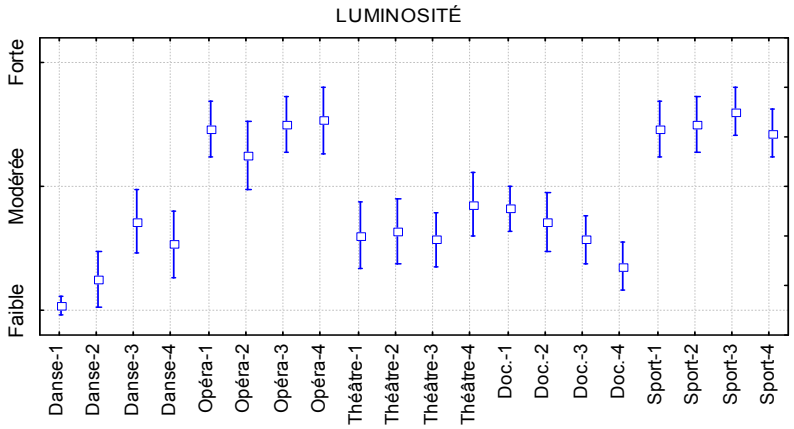


Fig. 7.3. Niveaux moyens obtenus pour le descripteur « Luminosité» de la catégorie Technique pour chaque séquence de test caractérisée par le panel naïf.

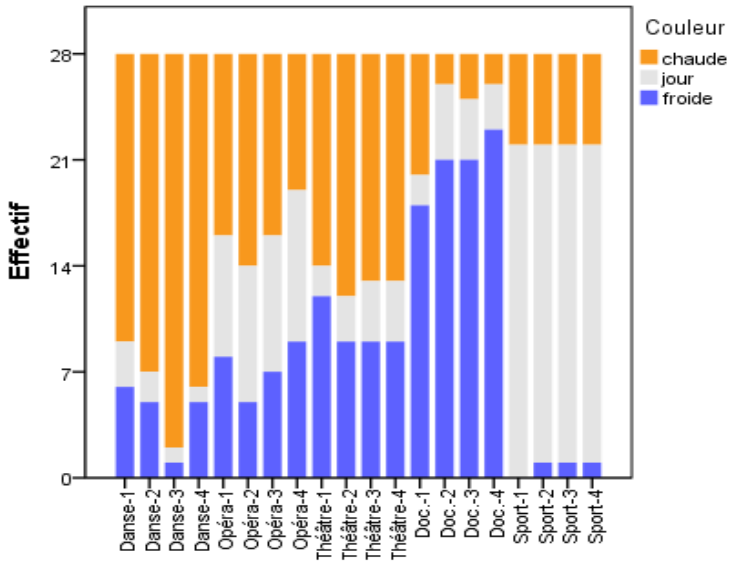


Fig. 7.4. Répartition des effectifs en fonction de la séquence pour l'annotation du descripteur « Couleur » de la catégorie Technique.

L'observation des figures indique clairement que les séquences des contenus *Opéra* et *Sport* ont été annotées comme les plus lumineuses du corpus. Par ailleurs, les séquences du contenu *Sport* ont été caractérisées par une température de couleur correspondant à « Jour »

(contenu tourné en extérieur) tandis que celles des contenus *Documentaire* et *Danse* ont principalement été définies respectivement par une couleur « Froide » et « Chaude ».

DESCRIPTEURS DE LA CATEGORIE HEDONIQUE

La Figure 7.5 ci-dessous présente les résultats obtenus (moyennes) pour les descripteurs *Intérêt*, *Valence* et *Arousal* de la catégorie Hédonique pour chaque séquence caractérisée par le panel naïf. Pour permettre la comparaison entre les descripteurs hédoniques évalués à partir d'échelles présentant différents niveaux (3 niveaux pour Intérêt et 9 niveaux pour Plaisir et Arousal), les données ont été normalisées entre 0 et 1 où 1 représente un score moyen élevé (Intérêt « Fort » par exemple) et 0 un score moyen faible.

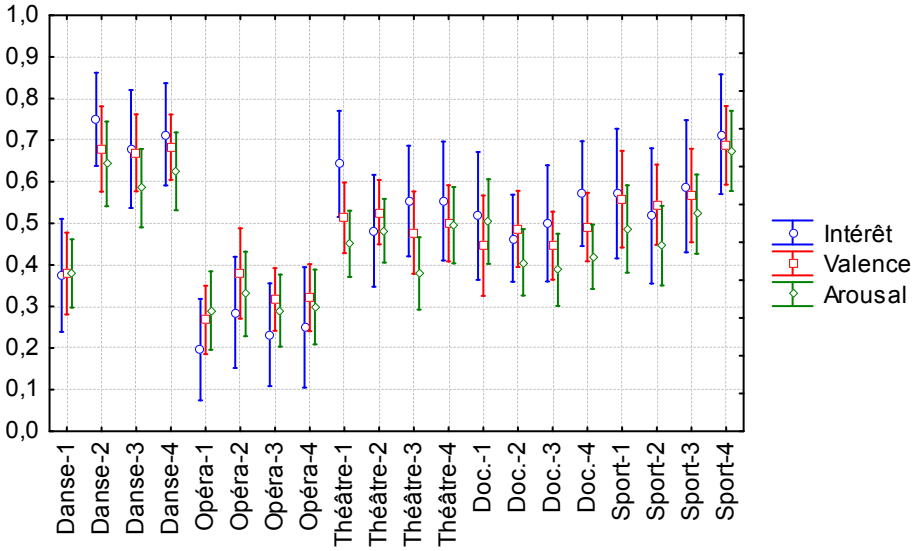


Fig. 7.5. Niveaux moyens obtenus pour les descripteurs « Intérêt », « Valence » et « Arousal » de la catégorie Hédonique pour chaque séquence de test caractérisée par le panel naïf où 1 représente un niveau élevé d'intérêt, de plaisir ou d'arousal et 0 un niveau faible d'intérêt, de plaisir ou d'arousal.

Un premier constat, suite à l'observation de la Figure 7.5, a trait aux séquences du contenu *Opéra*. Celles-ci se démarquent clairement du corpus de test par le faible niveau d'intérêt, de valence et d'arousal qu'elles ont suscité chez le participant. Ce constat tend à refléter une *qualité d'expérience* (envisagée sous l'angle de ces trois descripteurs) négative lors de la visualisation des séquences de ce contenu. A l'inverse, les séquences du contenu *Danse* ont reçu les notes les plus hautes pour ces mêmes descripteurs (excepté la séquence Danse-1).

Par ailleurs, les niveaux d'intérêt, de valence et d'arousal tendent à évoluer de manière similaire. Ainsi, lorsque le niveau d'un des trois descripteurs diminue alors le niveau des deux autres diminue également. Cependant, les participants ont été capables de distinguer les notions d'intérêt, de plaisir et d'arousal et de leur attribuer des notes sensiblement différentes pour une séquence donnée comme cela peut être observé pour les séquences Danse-3, Théâtre-1, Théâtre-3, Doc.-4, *etc.* Il semblerait que les descripteurs de la catégorie hédonique présentent une certaine complémentarité.

DESCRIPTEURS DE LA CATEGORIE SEMANTIQUE

Les figures ci-dessous présentent les niveaux moyens des descripteurs *Quantité d'information* (fig.7.6), *Compréhension* (fig.7.7), et *Dynamique de contenu* (fig.7.8), ainsi que la répartition des effectifs pour l'évaluation du descripteur *Modalité* (fig.7.9), obtenus pour chaque séquence.

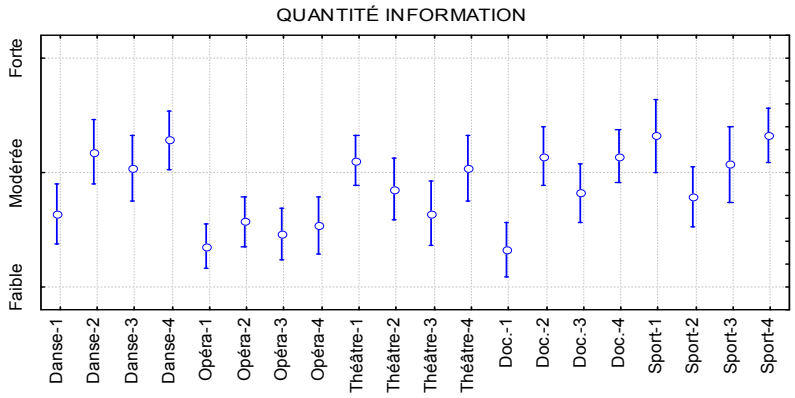


Fig. 7.6. Niveaux moyens obtenus pour le descripteur « Quantité d'information » (Quant. info) de la catégorie Sémantique pour chaque séquence de test caractérisée par le panel naïf.

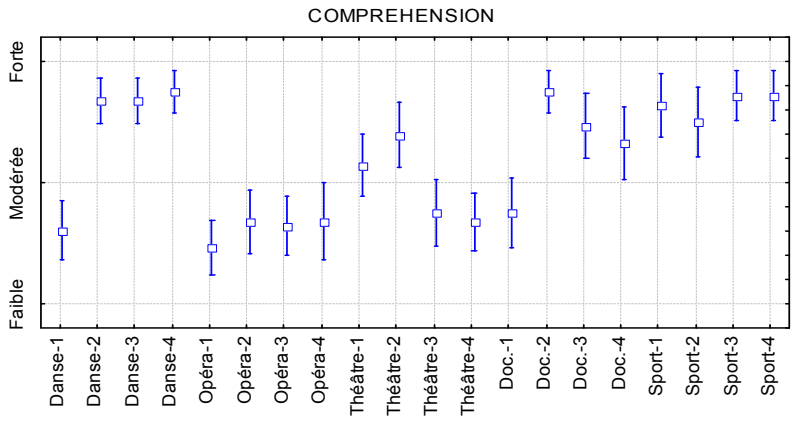


Fig. 7.7. Niveaux moyens obtenus pour le descripteur « Compréhension » de la catégorie Sémantique pour chaque séquence de test caractérisée par le panel naïf.

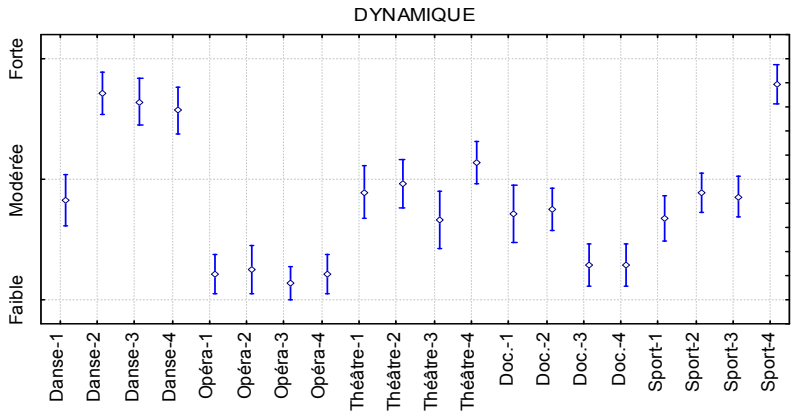


Fig. 7.8. Niveaux moyens obtenus pour le descripteur « Dynamique de contenu » de la catégorie Sémantique pour chaque séquence de test caractérisée par le panel naïf.

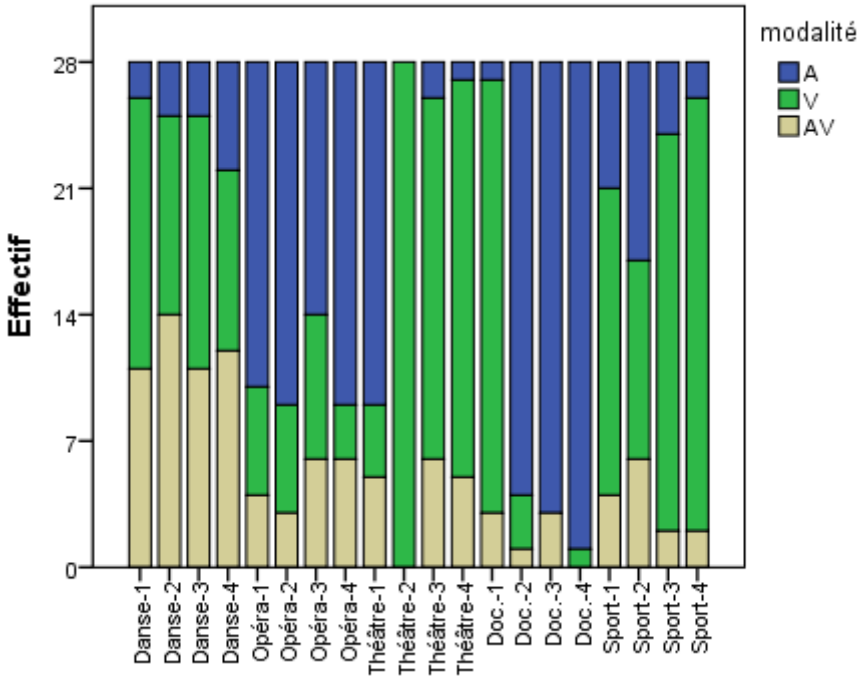


Fig. 7.9. Répartition des effectifs en fonction de la séquence pour l'annotation du descripteur « Modalité » de la catégorie Sémantique.

La Figure 7.9 indique que la totalité des séquences du contenu *Sport* et la majorité des séquences du contenu *Théâtre* ont été très largement considérées avec une dominance de la modalité vidéo. A l'inverse, la totalité des séquences du contenu *Opéra* et la majorité des séquences du contenu *Documentaire* ont été caractérisées par une modalité dominante audio. Il est intéressant de noter que la plupart des séquences à dominance vidéo ou audiovisuelle était accompagnée par une dynamique principalement jugée modérée voire forte. A l'inverse, aucune des séquences à dominance audio (Opéra, Doc.-2, Doc.-3 et Doc.-4) ne correspondait à une dynamique jugée «forte». Cette relation entre *Dynamique* et *Modalité* est confirmée par un Khi-deux d'indépendance de Pearson indiquant une association significative entre les deux

variables : χ^2= 75,45, ddl=4, $p < 0,01$. Le Tableau de contingence associé se trouve dans l'annexe 7-F. Ainsi, la notion de « Dynamique » est réservée au contenu vidéo, du moins dans le corpus de test utilisé ici.

Par ailleurs, les Figures 7.6 et 7.7 indiquent que les niveaux moyens de *Quantité d'information* et de *Compréhension* rapportés tendent à évoluer de la même manière que les notes de dynamique et que l'évaluations de l'expérience hédonique notamment concernant les séquences des contenus *Danse* et *Opéra*.

EVALUATION DE QUALITE

Malgré l'absence de dégradations liées à la transmission et aux conditions de restitution du flux audiovisuel (présentation des séquences en qualité full HD, 1080p), les participants ont perçu des différences de qualité A, V et AV entre les séquences (voir fig. 7.10 ci-dessous, voir annexe 7-E pour les effets significatifs). Ces différences de qualité, pouvant aller jusqu'à 7 points d'écart pour un individu donné, peuvent être attribuées au fait que les participants ont essayé de distribuer leurs jugements sur l'ensemble de l'échelle proposée. Néanmoins, les scores de qualité pourraient exprimer des différences de qualité perçues, en lien avec des différences techniques, sémantiques et/ou hédoniques. Les analyses *post-hoc* ont été effectuées à l'aide de tests HSD de Tukey.

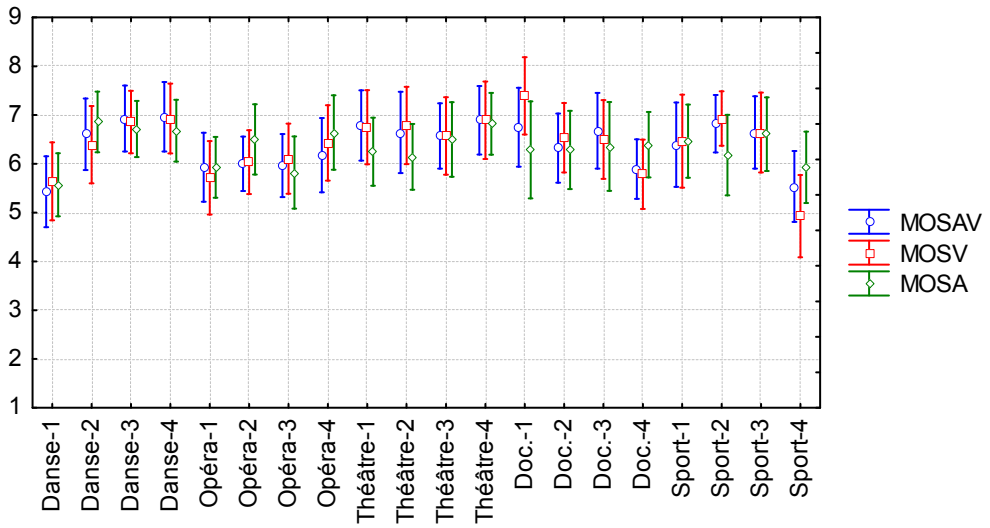


Fig. 7.10. MOSAV, V et A obtenues pour chaque séquence de test.

Plus précisément, Danse-1 a été jugée comment étant de qualité (AV et V) significativement inférieure à la séquence Danse-4 (p <0,001 pour QAV et p <0,05 pour QV). De la même manière, Sport-4 a reçu des notes de QAV et QV significativement plus basses que Sport-2 (p <0,01 pour QAV et p <0,001 pour QV). Ces différences pourraient être expliquées par des différences notables des catégories techniques, sémantiques et/ou hédoniques au sein d'un même contenu.

Le niveau de dynamique pourrait permettre d'expliquer les différences de qualité perçue constatées. En effet, Danse-1 comme Sport-4 se distinguaient des autres séquences de leur contenu d'appartenance par leurs niveaux de dynamique. Rappelons que les descripteurs techniques (*luminosité*, *couleur*, *dynamique caméra* et *détail)* ne variaient pas entre les séquences sélectionnées pour un contenu donné.

Sport-4 correspondait à l'unique séquence, de l'ensemble du corpus de test, combinant un niveau élevé de *Dynamique caméra* (caractérisation experte) et de *Dynamique du contenu*. Cette séquence offrait donc au participant un contenu hautement dynamique, c'est-à-dire riche du point de vue informationnel (grand nombre d'informations visuelles). Il est probable que le cumul des dynamiques (caméra et contenu) soit à l'origine d'une diminution de la qualité vidéo (considérée comme dominante pour ces séquences), l'audio ne présentant pas de différence significative avec les notes obtenues pour les autres séquences du contenu. Néanmoins, les évaluations des descripteurs hédoniques étaient élevées pour Sport-4. Cette séquence a en effet obtenu les moyennes les plus élevées, toutes séquences confondues, pour les descripteurs *Intérêt* (après la séquence Danse-2), *Valence* et *Arousal* traduisant une expérience hédonique pouvant être qualifiée de fortement positive.

A l'inverse, Danse-1 a été caractérisé par une modalité vidéo et un niveau faible de dynamique de contenu. Elle a également été qualifiée par des niveaux faibles ou modérés pour l'ensemble des descripteurs de haut-niveau *Intérêt* (*modéré)*, *Plaisir* (*3)*, *Arousal* (*5)*, *Quantité d'information* (faible) et *Compréhension* (faible). Danse-1 correspond donc à une séquence peu dynamique (notamment par rapport aux autres séquences du contenu, fortement dynamiques), pauvre en information tant auditive que visuelle et à l'origine d'une expérience spectateur plutôt négative. Il semblerait que plus la dynamique augmente, plus l'expérience hédonique soit positive et inversement. Le lien entre dynamique et expérience « hédonique » tend à être confirmé par les résultats obtenus pour l'ensemble des séquences du contenu *Opéra*. En effet, ces dernières ont toutes été caractérisées par un niveau faible de *Dynamique de contenu* (voir fig. 7.8 ci-dessus) et jugées par les participants avec un niveau faible d'intérêt, d'arousal et de valence (fig. 7.5).

Ainsi, la perception du niveau de dynamique tend à expliquer les différences de qualité perçues entre les séquences du contenu *Danse* et du contenu *Sport*. Dans le cas de Danse-1, l'absence de dynamique se traduirait par des notes de qualité moins bonnes. A l'inverse, une trop forte dynamique (cumul de dynamique caméra et contenu) comme c'est le cas pour Sport-4, diminuerait les niveaux de qualité vidéo et audiovisuelle tel que reflété par les notes de qualité.

Par ailleurs, un niveau faible de dynamique a été associé à des niveaux faibles d'intérêt, de plaisir et d'arousal (expérience hédonique négative). A l'inverse, une dynamique élevée a été à l'origine d'une expérience hédonique positive. Ainsi la notion de « Dynamique » participerait fortement à l'expérience « hédonique » du spectateur.

QUALITE D'EXPERIENCE

Les descripteurs des catégories Hédonique, Sémantique et Technique évalués dans cette étude constituent des facteurs potentiellement capables d'influencer la *qualité d'expérience* du participant. Afin d'étudier les descripteurs déterminant une expérience positive, une analyse de régression multiple simultanée a été conduite à partir de la caractérisation naïve obtenue (à savoir les moyennes obtenues pour chaque descripteur et chaque séquence) en considérant la variable dépendante « Valence » et les variables explicatives : « Intérêt», «Arousal», «Compréhension», «Dynamique», «Quantité d'information», « Luminosité » et « Qualité » (QAV, QV, QA) (les descripteurs « Modalité » et « Couleur » n'ont pu être intégrées en raison de leur nature nominale). L'analyse a révélé la participation de trois descripteurs : *Intérêt*, *Compréhension* et *Dynamique*. La valence de l'expérience peut être modélisée, dans le cadre de cette expérimentation, par l'équation suivante ($R=0,99$, $R^2 = 0,98$) :

$$\text{Valence de l'expérience} = 1,37\times\text{Intérêt} + 0,56\times\text{Compréhension} + 0,40\times\text{Dynamique} - 0,76$$

Ce résultat indique que l'augmentation des niveaux d'intérêt, de compréhension et de dynamique participe à la valence positive de l'expérience du spectateur. La *qualité d'expérience* du spectateur, envisagée sous l'angle de sa valence semble donc pouvoir être prédite par un sous-ensemble d'indicateurs.

7.2.9.3. CARACTERISATION FINALE DES SEQUENCES

Les précédents paragraphes ont présenté la caractérisation naïve des séquences (annexe 7-D). Cependant, la description des contenus à partir des descripteurs de bas-niveau sera toujours réalisée à partir de la caractérisation experte. Le Tableau 7.5 ci-dessous présente la description obtenue pour chacune des vingt séquences de l'expérimentation B1 à savoir les descripteurs de haut-niveau des catégories *Hédonique* et *Sémantique* (caractérisation naïve) et de bas-niveau des catégories *Sémantique* et *Technique* (caractérisation experte).

Tableau 7.5. Tableau présentant la caractérisation des séquences réalisée par l'expert et par le panel de spectateur « naïfs ».

	C. Expert									C. Naïve				
Séq.	**Mod**	**R.AV**	**EX.S**	**Nb P**	**D.Ct**	**Lum**	**Dét**	**D.Cm**	**Col**	**Int**	**Val**	**Ar**	**Info**	**Comp**
Danse-1	V	In	MuS	F	Fa	Fa	M	Fa	C	M	3	5	Fa	Fa
Danse-2	V	In	MuS	F	F	Fa	M	Fa	C	F	7	7	M	F
Danse-3	V	In	MuS	F	F	Fa	M	Fa	C	F	7	6	M	F
Danse-4	V	In	MuS	M	F	Fa	M	Fa	C	F	7	7	M	F
Opéra-1	A	HC	P	F	Fa	M	M	Fa	J	Fa	3	2	Fa	Fa
Opéra-2	A	In	P	F	Fa	M	M	Fa	J	Fa	4	4	M	M
Opéra-3	A	In	P	F	Fa	M	M	Fa	J	Fa	3	3	Fa	M
Opéra-4	A	In	P	Fa	Fa	M	M	Fa	J	Fa	4	3	Fa	Fa
Théâtre-1	A	In	P	Fa	Fa	Fa	M	Fa	F	M	6	5	M	M
Théâtre-2	V	In	B	Fa	Fa	Fa	M	Fa	F	M	5	5	M	F
Théâtre-3	V	Off	MuS	M	M	Fa	M	Fa	F	M	5	3	Fa	M
Théâtre-4	AV	In	B	M	M	Fa	M	Fa	F	M	5	5	M	M
Doc-1	V	Off	MuS	Fa	M	Fa	M	Fa	J	M	3	5	Fa	Fa
Doc-2	A	Off	P	Fa	M	Fa	M	Fa	J	M	5	3	M	F
Doc-3	A	In	P	Fa	Fa	Fa	M	Fa	J	M	5	5	M	F
Doc-4	A	Off	P	Fa	Fa	Fa	M	Fa	J	M	5	5	M	F
Sport-1	AV	Off	P	F	Fa	F	F	F	J	F	7	5	F	F
Sport-2	A	Off	P	F	Fa	F	F	F	J	F	7	6	M	F
Sport-3	V	Off	P	F	M	F	F	F	J	F	6	5	F	F
Sport-4	V	Off	P	F	F	F	F	F	J	F	7	7	M	F

Caractérisation des séquences (Séq.) réalisée par l'expert (C.Expert) à partir des descripteurs de bas-niveau Sémantique : Modalité (Mod), Relation AV (R.AV), Expression Sonore (EX.S), Nombre de personnages (Nb P), Dynamique de Contenu (D.Ct) et Technique : Luminosité (Lum), Détail (Dét), Dynamique caméra (D.Cm) et Couleur (Col) ; selon les niveaux Faible (Fa), Modéré (M), Fort (F) ou Audio (A), Vidéo (V), AudioVisuel (AV) ou Chaude (C), Jour (J), Froide ou Musique (Mus), Parole (P), Bruit (B) et par le panel de spectateur « naïfs » (C.Naïve) selon le calcul du mode réalisé pour chaque séquences et chaque descripteur de haut-niveau des catégories Hédonique : Intérêt (Int) Valence (Val) et Arousal (Ar) et Sémantique : Quantité d'information (Info) et Compréhension (Comp).

7.3. CONCLUSIONS B1

Un des principaux objectifs de l'expérimentation B1 était de caractériser les contenus de tests dans l'intention de mieux comprendre la relation entre contenu et la qualité perçue. Cette caractérisation a été réalisée par un expert et a permis de **disposer d'un ensemble de contenus pour lequel les caractéristiques techniques et sémantiques de bas-niveau sont connues**. Dans les études suivantes, chaque séquence extraite d'un de ces cinq contenus pourra donc être définie sur la base de ces descripteurs.

Dans un second temps, la pertinence d'un sous-ensemble de descripteurs (considérés comme étant les plus subjectifs) utilisé par l'expert a été vérifiée auprès de participants naïfs.

La **concordance entre la caractérisation de l'expert et celle des participants** pour les descripteurs ***Modalité*** **et** ***Dynamique de contenu*** **confirme la pertinence de ces critères**. Le descripteur de luminosité a rencontré un plus faible accord entre expert et naïfs, principalement concernant la qualification des séquences par le niveau « faible ». Malgré le désaccord observé pour ce niveau, le terme semble tout de même avoir été correctement assimilé par les naïfs. En revanche, les termes utilisés pour qualifier **la température de couleur semblent réservés au domaine technique de l'audiovisuel** et peu accessibles à un panel de participants non experts. Ce descripteur, peu compréhensible pour des participants non experts, ne sera pas être intégré à un futur questionnaire d'évaluation.

Cette étude a également permis de disposer d'un corpus de séquences courtes (8 à 10 s) caractérisées non seulement par des critères techniques et sémantiques de bas-niveau mais également par des critères hédoniques et sémantiques de haut-niveau. Globalement, les résultats ont montré que **les participants ont été en mesure de décrire les séquences visualisées à partir des descripteurs proposés.** Chaque descripteur semble apporter des informations pertinentes sur les plans techniques, sémantiques ou hédoniques pour décrire les contenus audiovisuels visualisés.

Les évaluations des descripteurs par les participants ont donc été influencées par la séquence mais aussi, plus largement, par le contenu. Par exemple, *Opéra* a suscité peu d'intérêt ainsi qu'une valence négative et un arousal faible chez le participant. Cela pourrait s'expliquer par le fait que les séquences choisies ont été extraites d'un contexte plus général, c'est-à-dire qu'elles ont été désolidarisées d'une trame globale porteuse de sens. Ces bribes d'évènements auraient pu couper le participant du sens général du contenu. Ceci est d'autant plus vrai que les séquences du contenu *Opéra* présentent un contexte de langue étrangère (le sens est donc d'autant plus difficile à extraire, limitant la compréhension de l'extrait) pour des séquences dont la modalité dominante a été jugée comme étant l'audio. Ainsi, il conviendrait, afin d'éviter ce désintérêt, de **replacer le participant dans la trame narrative globale du contenu** par exemple en proposant la lecture de synopsis avant l'évaluation des séquences tests.

Les résultats ont aussi mis en avant un **lien entre les descripteurs** ***Dynamique*** **et** ***Modalité dominante***. En effet, la dynamique a été associée, dans le cadre du corpus de séquences de test proposé ici, à la modalité vidéo. Précisément, une prédominance de l'audio serait principalement associée à une dynamique faible et une prédominance de la vidéo serait majoritairement assimilée à une dynamique modérée voire forte. Ce constat va dans le sens de celui apporté par Hands (2004) supposant que la qualité audio serait dominante pour un contexte AV faiblement dynamique tandis que la qualité vidéo serait dominante pour un contexte AV fortement dynamique.

Malgré l'absence de dégradations, des différences de notations ont pu être observées concernant la qualité perçue. Il semblerait que le descripteur *Dynamique* soit un bon candidat pour expliquer ces différences. En effet, les séquences peu dynamiques (Danse-1) et

a fortiori audio (séquences du contenu *Opéra*), apportant peu d'informations au spectateur et à l'origine d'une expérience hédonique négative, ont été caractérisées par des niveaux faibles de qualité perçue. A l'inverse, une séquence hautement dynamique et *a fortiori* vidéo, apportant de nombreuses informations aux spectateurs (Sport-4) et à l'origine d'une expérience fortement positive (du point de vue de l'intérêt, du plaisir et du niveau d'arousal), a entraîné un niveau de qualité perçue altérée. Ce résultat indique également que les participants ont été capables, dans un contexte de qualité non dégradée, de juger de la qualité des signaux audio et/ou vidéo indépendamment de la tendance positive ou négative de leurs expériences.

Enfin, cette étude a également permis de mieux comprendre le lien entre le contenu (étudié à partir des descripteurs évalués) et la *qualité d'expérience* envisagée sous l'angle de sa valence. En effet, **un sous-ensemble de descripteurs (intérêt, compréhension et dynamique) permettrait de prédire la valence de l'expérience**. Ces descripteurs pourraient être pris en compte dans le cadre de méthodes portant sur l'évaluation de la qualité des signaux audio et vidéo restitués.

7.3.1. RETOUR SUR EXPERIMENTATION A : INTERPRETATIONS ET EXPLICATIONS

La caractérisation des contenus par l'expert permet déjà de confirmer ou tout au moins de consolider les éléments d'explications soumis lors de l'expérimentation A.

Une influence du niveau de luminosité sur les mesures du diamètre pupillaire (DP) était supposée (visualisation des contenus Documentaire, Opéra et Sport, pour rappel : DP Documentaire > DP Sport et Opéra). La caractérisation experte des séquences tend à confirmer ce postulat. En effet, les contenus *Opéra* et *Sport* ont été caractérisés comme les plus lumineux (d'après la caractérisation experte réalisée sur les contenus entiers, le niveau de luminosité variait peu au sein d'un même contenu). De ce fait, le diamètre pupillaire significativement plus petit pour ces deux contenus, par rapport au contenu *Documentaire*, peut s'expliquer par un effet du niveau de luminosité (confirmation d'un réflexe photo-moteur).

Un effet du niveau de dynamique a également été suggéré pour expliquer l'augmentation de l'activité électrodermale (AED) durant le contenu *Sport* par rapport aux contenus *Opéra* et *Documentaire*. La caractérisation experte des séquences tend également à consolider ce postulat : *Sport* a été caractérisé par un niveau de dynamique de contenu plus élevé que les contenus *Documentaire* et *Opéra*. Une autre différence notable entre ces contenus provient du niveau de dynamique de caméra. En effet, *Sport* présentait un niveau élevé pour ce descripteur tandis qu'à l'inverse, la dynamique caméra pour *Documentaire* et *Opéra* était faible ou modérée. **Ainsi, il semblerait que l'AED soit sensible aux aspects de dynamique relatifs à l'activité à la fois des personnages ou objets d'intérêt et des caméras.** Comme l'ont indiqué Lang A. *et al.* (1999), Lang A. *et al.* (2000), Simons *et al.* (1999) ou encore Yoon *et al.* (1998), l'AED est sensible (augmentation) au mouvement et aux

changements de plan ou de scène. La caractérisation tend à aller dans le sens des constats de ces auteurs et à indiquer que l'AED pourrait être sensible au niveau de dynamique de manière générale (action des personnages, changement de plan, zooms, travelling, *etc.*).

Ainsi, la caractérisation réalisée accrédite ou consolide les postulats émis pour expliquer les résultats de l'expérimentation A. Ce constat souligne l'importance de caractériser les contenus de test pour une meilleure compréhension de l'influence de la qualité sur la perception du spectateur et les mesures psychophysiologiques recueillies.

La caractérisation réalisée par l'expert puis par le panel de participants « naïfs » a donc permis d'apporter des éléments de réponses quant aux facteurs capables d'influencer la perception de qualité (niveau de dynamique par exemple) et la *qualité d'expérience* du spectateur subjectivement ou physiologiquement étudiée. Afin de mieux comprendre les interactions entre contenu et perception de qualité, le corpus de séquences de test précédemment utilisé et caractérisé a été présenté avec différentes conditions de qualité lors d'une seconde expérimentation (B2).

7.4. EXPERIMENTATION B2 : CONTENU ET QUALITE

7.4.1. OBJECTIFS

L'objectif de l'étude B2 est d'étudier l'influence du contenu à travers des séquences caractérisées lors de l'expérimentation B1 (c'est-à-dire décrites par un certain nombre de descripteurs) sur la perception de la qualité audiovisuelle. La connaissance des descripteurs les plus critiques du point de vue de l'évaluation de qualité doit permettre d'optimiser l'étape de caractérisation des séquences de test dans le cadre d'une méthode d'évaluation de la *qualité d'expérience* du spectateur.

7.4.2. PARTICIPANTS

Dans cette expérimentation, trente-quatre participants (16 femmes, 18 hommes) entre 21 et 60 ans devaient évaluer le niveau de qualité audiovisuelle après chaque séquence de test visualisée. Les participants étaient rémunérés pour leur participation.

7.4.3. MATERIEL

7.4.3.1. CONFIGURATION GENERALE

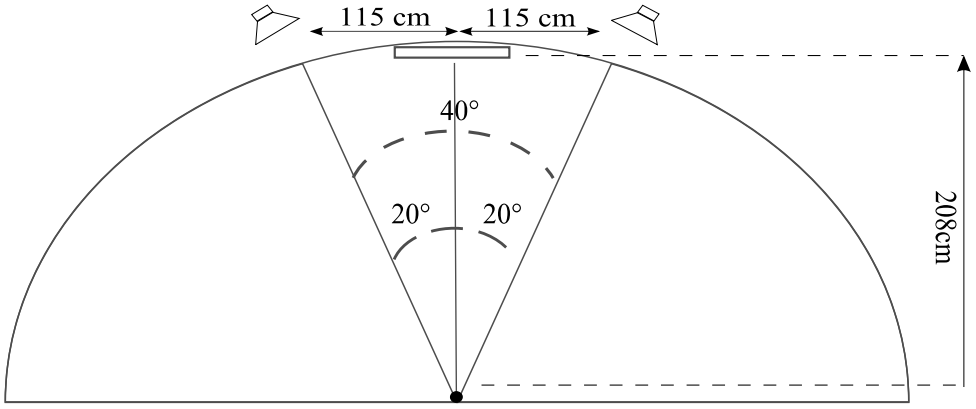


Fig. 7.11. Schéma de la configuration de la salle de test (250×310×320 cm) de l'expérimentation B2. La place du participant est représentée par un point noir, l'écran est figuré par un rectangle.

L'expérience a été conduite dans une salle insonorisée, les conditions de visualisation et d'écoute correspondaient aux recommandations de la norme UIT-T P.911 (voir annexe 7-G). Les paramètres de brillance et de contraste de l'écran respectaient la recommandation UIT-R BT.814-2 (UIT, 2007). Concernant l'affichage, un écran LCD de 46" (117 cm), full HD (1080p, 16:9) de type Sony modèle KDL-46HX920 a été utilisé. La distance de visualisation, en accord avec la norme UIT-T P.911, a été fixée à 3,2 fois la hauteur de l'écran, soit 192 cm.

Des haut-parleurs Genelec de modèle 8040A ont été réglés à une hauteur de 98 cm et placés de manière à être équidistants du centre de l'écran (115 cm) et de la tête du participant (208 cm). La Figure 7.11 ci-dessus présente la configuration de la salle de test, celle-ci respectait les recommandations fournies par la norme UIT-R BS.1286 (UIT, 1997). En accord avec la recommandation UIT-T P.911, le volume sonore restitué par les HP se situait autour de 80 dB (mesuré à l'aide d'un sonomètre), le son ambiant de la salle de test était inférieur à 31 dB.

7.4.3.2. CONFIGURATION TECHNIQUE

Comme l'indique la Figure 7.12 ci-après, le signal audiovisuel global était stocké et diffusé par un magnétoscope numérique (DVS - Digital Video System - Pronto2K). Les signaux audio et vidéo étaient ensuite acheminés séparément vers le matériel de restitution audio (c.-à-d. HP externes *via* un amplificateur -SPL 2380-) ou vidéo (écran). Avant cela, chaque signal était converti au bon format d'entrée, c'est-à-dire d'un signal numérique (signal de sortie du DVS) à un signal analogique (signal d'entrée HP) pour l'audio (convertisseur Méridian 566) et une conversion SDI (signal de sortie du DVS) vers DVI (signal de sortie du convertisseur, Gefen HD-SDI to DVI scaler) pour permettre une entrée HDMI (signal d'entrée écran) pour la vidéo.

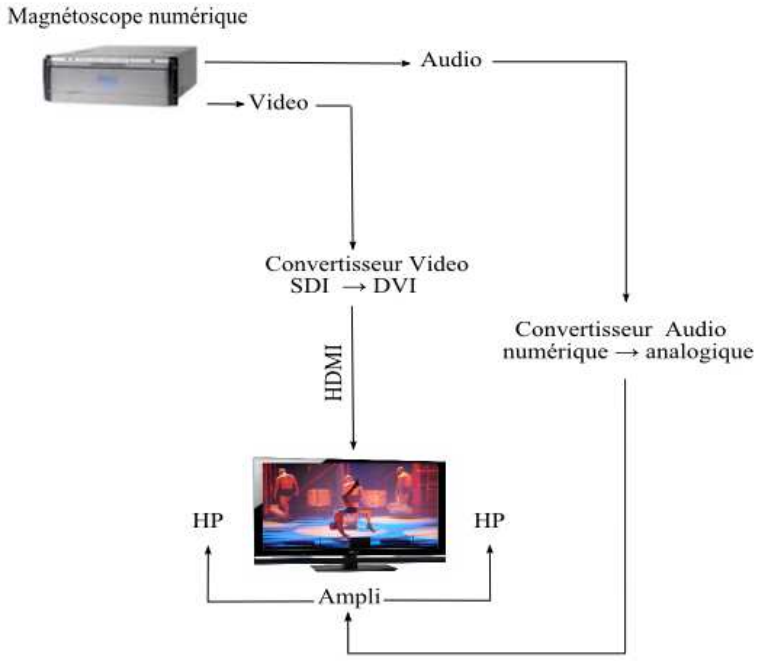


Fig. 7.12. Configuration technique de l'expérimentation B2.

Le logiciel SEOVQ (Subjective Evaluation and Optimization of Video Quality, solution d'interface pour protocole d'évaluation d'images multimédias) a permis de déclencher, à partir de l'interface utilisée, la lecture des séquences vidéo (les séquences ne pouvaient vues et entendues qu'une seule et unique fois). Cette configuration permettait l'utilisation de *playlist*, avec un ordre aléatoire de présentation des séquences différent pour chaque participant.

7.4.4. STIMULI

Les vingt séquences de test utilisées lors de l'expérimentation B1 ont été présentées en format 2D full HD 1080p avec dix conditions de qualité différentes choisies pour être représentatives de dégradations susceptibles de survenir dans des conditions réelles de visualisation. L'ensemble des séquences audio a été normalisé en niveau sonore. Les conditions de qualité étaient les suivantes :

- **Condition1 (REF)** : séquence audiovisuelle de référence (sans dégradations) présentant un signal audio de 48 Kbps/16 bit et un signal vidéo .avi non compressé,
- **Condition 2 (D)** : séquence audiovisuelle présentant une désynchronisation entre l'image et le son avec un délai de 1500 ms appliqué sur le signal vidéo seulement (retard de l'image par rapport au son). Les caractéristiques des signaux audio et vidéo étaient identiques à celles de la séquence référence,
- **Condition 3 (V-DEB)** : séquence audiovisuelle présentant une réduction du débit vidéo. Sur la base des résultats de l'expérimentation A, un débit adaptatif a été

appliqué dans cette étude. Le signal vidéo était compressé avec le codeur AVC-x264 pour un débit de 93 à 1600 kbps selon le niveau de détail (complexité) de la séquence à coder. L'utilisation d'un débit adaptatif permet d'obtenir des dégradations plus homogènes du point de vue perceptif entre les différentes séquences (d'après les résultats de l'expérimentation A un seuil identique de débit appliqué à un contenu vidéo « simple » - plus facile à coder - dégraderait moins la qualité perçue qu'un contenu complexe),

- **Condition 4 (V-Gel) :** séquence audiovisuelle présentant un gel vidéo pour laquelle une série d'image "gelées" (figées) était introduites. Plus précisément, cinq périodes de "gel" d'une durée de 18 images chacune (soit 720 ms pour un total de 3,6 s dégradées) étaient introduites de manière aléatoire sur les dix secondes présentées,
- **Condition 5 (A-PP)** : séquence audiovisuelle présentant une perte de paquets d'information audio pour laquelle un taux de 10% de perte de paquets était introduit aléatoirement sur le flux audio,
- **Condition 6 (A-DEB)** : séquence audiovisuelle présentant une réduction du débit audio (A-DEB) avec le signal audio compressé à 64Kbps/8Ko (codeur AC3),
- **Condition 7** : séquence audiovisuelle présentant la combinaison des dégradations A-PP * V-Gel,
- **Condition 8** : séquence audiovisuelle présentant la combinaison des dégradations A-PP * V-DEB,
- **Condition 9** : séquence audiovisuelle présentant la combinaison des dégradations A-DEB + V- DEB,
- **Condition 10** : séquence audiovisuelle présentant la combinaison des dégradations A-DEB + V- Gel.

7.4.5. PROTOCOLE

Chaque participant visualisait un total de deux cents séquences (20 séquences × 10 conditions de qualité) d'une durée de huit à dix secondes, présentées dans un ordre aléatoire différent pour chaque individu. Entre chaque séquence, les participants disposaient d'une pause de cinq secondes pour évaluer la qualité AV globale sur une échelle en neuf points et cinq items (*Excellent*, *Bon*, *Satisfaisant*, *Médiocre*, *Mauvais*) conformément à la méthode ACR de la recommandation UIT-T P.911. Chaque séquence était visualisée une seule et unique fois. Le logiciel SEOVQ a été utilisé pour recueillir les jugements des participants. Une illustration de l'interface utilisée est apportée par la Figure 7.13 ci-dessous. Avant la passation, une fiche de consignes était distribuée au participant. Les consignes sont présentées dans l'annexe 7-H. La durée totale de la passation de test était d'environ une heure trente.

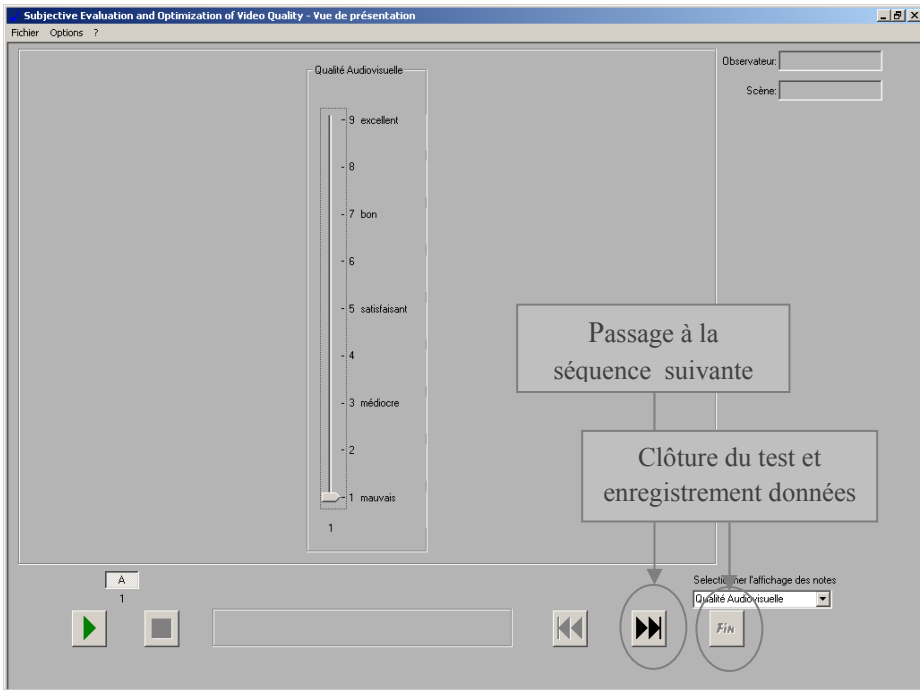


Fig. 7.13. Interface d'évaluation du logiciel SEOVQ permettant au participant de reporter son jugement sur l'échelle affichée. Une fois l'évaluation effectuée, le participant passait à la séquence suivante au moyen d'un *clic bouton* (a). Chaque séquence ne pouvait cependant être visualisée et entendue qu'une seule et unique fois. Les résultats pour chaque participant étaient ensuite enregistrés (b), après chaque passation, dans un fichier dédié.

7.4.6. OBSERVABLES ET HYPOTHESES

Dans cette étude, seule la qualité audiovisuelle globale était évaluée par les participants, conformément à la norme UIT-T P.911. L'objectif de cette expérimentation était d'étudier l'influence du contenu sur la qualité audiovisuelle perçue. En conséquence, il était attendu que les notes de qualité audiovisuelle (MOSAV) recueillies ne dépendent pas seulement des dégradations de qualité mais aussi du contenu qualifié par les descripteurs de bas et haut-niveau précédemment définis.

H0 : au-delà d'un effet de la dégradation, les scores MOSAV sont également influencés par la séquence de test (décrite à partir des descripteurs de B1).

7.4.7. RESULTATS

Une ANOVA à mesures répétées considérant la variable indépendante « Séquences » et la variable dépendante « Qualité » à dix modalités a indiqué un effet significatif de ces deux variables sur les scores MOSAV obtenus, à savoir F (19, 627) = 377,60, p <0,001 pour l'effet de la séquence et F (9, 297) = 15,61, p <0,001 pour l'effet de la qualité. Une interaction de la variable « Séquences » et de la variable « Qualité » a également été constatée : F (171, 5643) = 12,33, p <0,001. Ces résultats confirment H0 en montrant que les séquences de test (décrites à partir des descripteurs de B1) influencent la perception de qualité. Les notes

MOSAV obtenues pour chaque dégradation sont présentées pour chaque contenu par les figures ci-dessous : Danse (fig. 7.14), Documentaire (fig. 7.15), Opéra (fig. 7.16), Théâtre (fig.7.17) et Sport (fig. 7.18). Une présentation par type de dégradation est également apportée dans l'annexe 7-I. Cette présentation permet de mieux observer certains effets discutés lors de l'analyse des résultats.

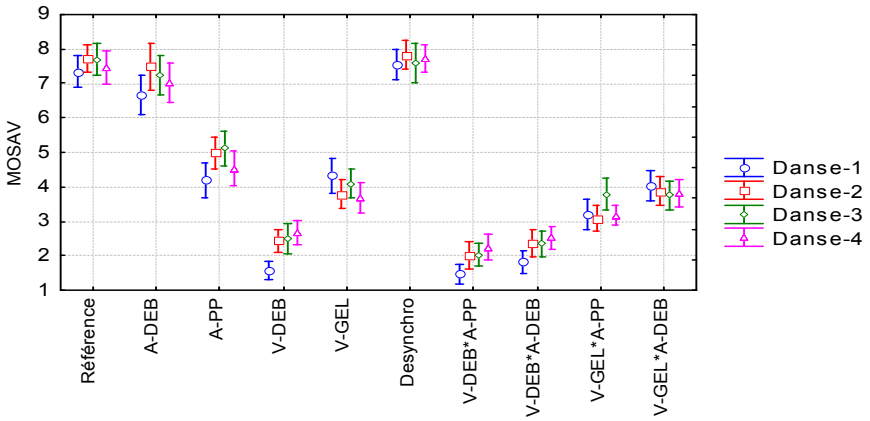


Fig. 7.14. MOSAV obtenues pour chaque condition de qualité pour les séquences du contenu Danse.

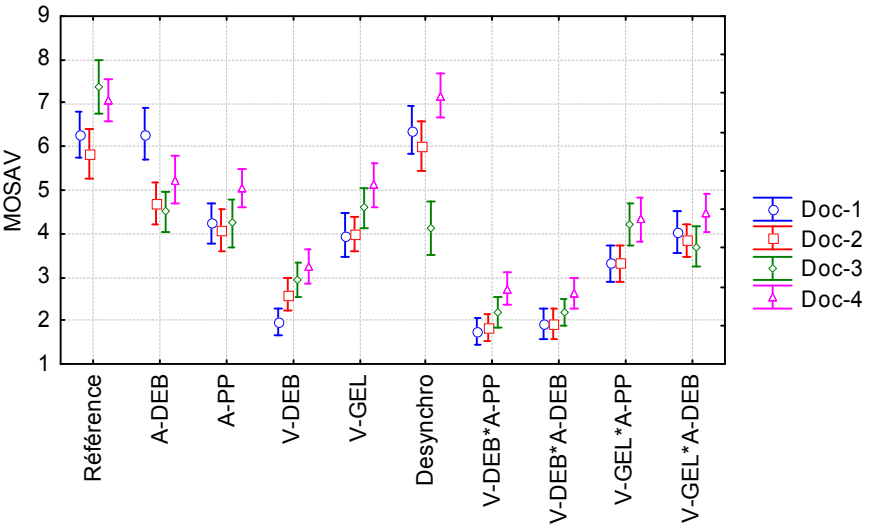


Fig. 7.15. MOSAV obtenues pour chaque condition de qualité pour les séquences du contenu Documentaire (Doc).

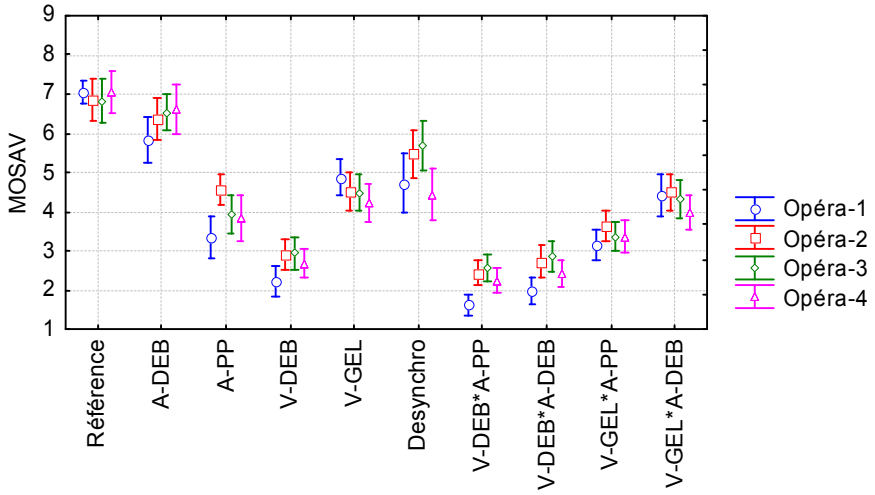


Fig. 7.16. MOSAV obtenues pour chaque condition de qualité pour les séquences du contenu Opéra.

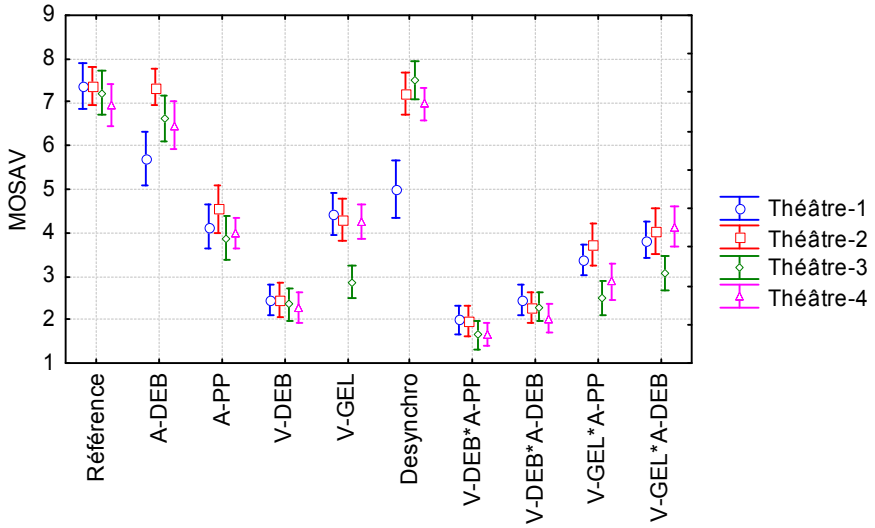


Fig. 7.17. MOSAV obtenues pour chaque condition de qualité pour les séquences du contenu Théâtre.

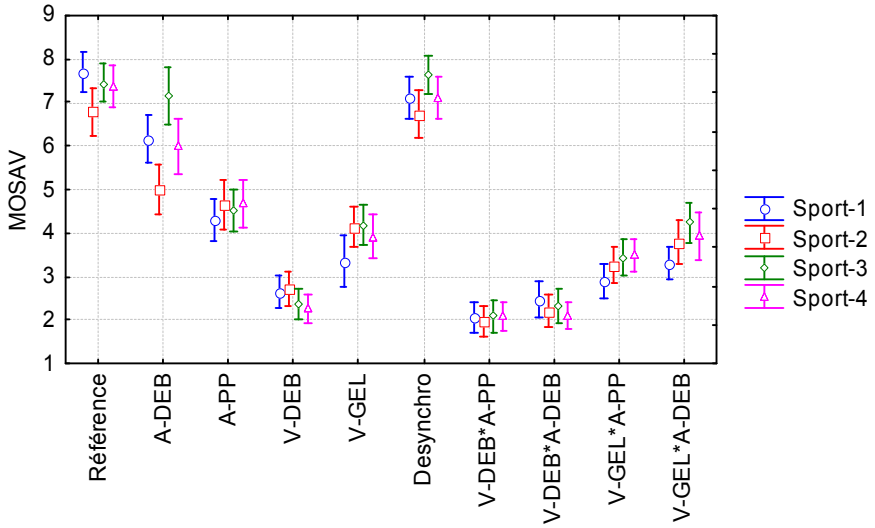


Fig. 7.18. MOSAV obtenues pour chaque condition de qualité pour les séquences du contenu Sport.

7.4.7.1. EFFET DE LA DESYNCHRONISATION

Les figures ci-dessus indiquent que la désynchronisation a été à l'origine d'une diminution des notes MOSAV sur l'ensemble des séquences extraites du contenu *Opéra* ainsi que sur les séquences Théâtre-1 et Doc-3 (diminution de la note d'environ deux points par rapport aux séquences des autres contenus). Ces séquences ont en commun un contexte audio à la fois verbal et diégétique. La totalité des autres séquences, pour lesquelles la désynchronisation ne semble pas avoir dégradé la qualité, correspondait soit à des scènes auditives caractérisées par les modes *Musique* ou *Bruit* du descripteur « Expression sonore », que ce dernier soit caractérisé comme diégétique (Danse, Théâtre-2 et -4) ou non (Théâtre-3, Doc.-1), soit à des contenus verbaux (mode Parole) mais présentant un son extra-diégétique (son *off*) comme c'est le cas pour le contenu *Sport* et les séquences Doc-4, Doc-2 (voir annexe 7-B). Ainsi, pour altérer la perception de qualité, la désynchronisation ne doit pas seulement survenir lors de séquences verbales mais lors de séquences verbales présentant aussi un son diégétique (*in* et hors-champ).

7.4.7.2. EFFET DES DEGRADATIONS AUDIO

Les figures ci-avant montrent que globalement la dégradation A-PP a été perçue comme dégradant plus fortement la qualité que la dégradation A-DEB. Par ailleurs, le niveau de qualité perçue de certaines séquences semble avoir été plus fortement influencé par les dégradations audio. Il s'agit notamment des séquences : Opéra-1 (plus fortement dégradée par A-PP que les autres séquences du contenu), Théâtre-1 et Sport-2 (plus fortement dégradées par A-DEB que les autres séquences du contenu) et les séquences Doc-2, Doc-3 et Doc-4 du contenu *Documentaire* (plus fortement dégradées par A-DEB que l'ensemble des séquences du corpus, exceptée Sport-2). La connaissance de la modalité dominante permet de mieux comprendre ces différences. En effet, neuf séquences ont été caractérisées par une modalité dominante audio : Opéra (1, 2, 3 et 4), Théâtre 1, Documentaire (2, 3 et 4) et Sport-2 (voir Annexe 7-B). Ainsi, il semble que la modalité dominante permette d'expliquer les différences observées. Le fait que les séquences du contenu *Opéra* n'aient pas été plus fortement dégradées par A-PP et A-DEB que les autres contenus (en-dehors d'Opéra-1) pourrait s'expliquer par le fait que les scènes auditives correspondaient à des paroles chantées et de langue étrangère. Il est alors envisageable que les dégradations audio dégradent l'intelligibilité du contenu sonore, celle-ci n'étant pas dégradée dans le cas d'un contenu de paroles étrangères (pas d'accès au sens).

L'influence plus importante de la dégradation audio lors de contenus à dominante audio semble notamment vraie pour la dégradation A-DEB. Le fait que la dégradation A-PP ait moins été soumise à l'influence du contenu pourrait s'expliquer par un effet moins subtil de cette dégradation qui viendrait lisser les effets de contenu (dégradation plus fortement perceptible et ce, indépendamment du contenu).

7.4.7.3. EFFET DES DEGRADATIONS VIDEO

Globalement la dégradation V-DEB a été perçue comme dégradant plus fortement la qualité que V-GEL. Pour cette dernière dégradation, les notes de QAV attribuées aux séquences des contenus *Opéra*, *Documentaire* et *Théâtre* tendent à être plus élevées que celles attribuées aux séquences des contenus *Danse* et *Sport* (pour une meilleure observation de cette tendance voir annexe 7-I). Comme indiqué dans le paragraphe ci-dessus, la modalité dominante constitue un élément d'explication des variations observées : les séquences des contenus *Opéra* et *Documentaire* étaient principalement *Audio* tandis que celles des contenus *Danse, Sport* et *Théâtre* étaient principalement *Vidéo* ou *Audiovisuelle*. Néanmoins, le descripteur de modalité ne permet pas d'expliquer les notes attribuées aux séquences du contenu *Théâtre*. Une influence du niveau de dynamique peut ici être supposée : toutes les séquences des contenus *Sport* et *Danse* présentaient une dynamique de contenu élevée (Danse, excepté Danse-1) et/ou une dynamique de caméra élevée (Sport), ce qui n'était pas le cas des séquences du contenu *Théâtre*. Ainsi, la présence de saccades (gel d'images) dégraderait plus fortement la qualité lorsque la séquence présente une dynamique forte (mouvements des personnages ou de caméra plus difficiles à suivre en présence de saccades).

Une autre remarque concerne la séquence Théâtre-3 ayant obtenu la note MOSAV la plus faible du corpus lors de la présentation de V-GEL. Cette séquence était la seule avec Doc-1 à présenter à la fois une modalité dominante vidéo et un son *off* de musique. En d'autres termes, aucune information auditive complémentaire ne pouvait permettre au participant de compenser la dégradation vidéo. Le fait que Doc-1 ait reçu des notes de QAV supérieures à Théâtre-3 pourrait être expliqué par le nombre de personnages. En effet, Doc-1 était caractérisée par un niveau faible (1 seul personnage) tandis que Théâtre-3 présentait un niveau modéré (3 personnages) dont le suivi des déplacements serait rendu plus difficile par la présence de saccades vidéo (images gelées). Ainsi, V-GEL aurait plus fortement dégradé Théâtre-3 en raison de l'absence d'information auditive pertinente (musique *off*) et en présence de mouvements sur le média vidéo (déplacements des personnages). Ce constat tend à appuyer l'impact plus important de V-GEL lorsque cette dégradation est appliquée lors de séquences dynamiques (mouvements des caméras ou des personnages) par rapport à des séquences peu dynamiques.

La note MOSAV la plus faible du corpus de séquences de test a été attribuée à la séquence Danse-1 en présence de la dégradation V-DEB. Cette séquence s'était également distinguée lors de l'expérimentation B1 (notes plus faibles de QAV et QV par rapport aux autres séquences du contenu) bien qu'aucune dégradation n'était présente. Danse-1 correspond à une séquence pauvre du point de vue informationnel tant sur le média audio (mode *Musique* du descripteur *Expression Sonore*) que vidéo (niveau *Faible* des descripteurs *Dynamique de contenu* et *Dynamique de caméra*). Danse-1 était la seule séquence du corpus à présenter la combinaison « musique » (modalité dominante vidéo) et « dynamique faible » (contenu et caméra). Le fait que la diminution de qualité observée pour Danse-1 ne soit pas retrouvée en présence de V-GEL laisse supposer qu'une dégradation par diminution du débit vidéo (perte

de résolution) dégrade plus fortement la qualité perçue, que la présence de saccades vidéo, pour ce type de séquence (modalité dominante vidéo -musique- et dynamique faible).

7.4.7.4. EFFET DES DEGRADATIONS AUDIO-VIDEO

De manière générale, les variations des notes MOSAV retranscrivent celles observées lors de l'analyse des dégradations vidéo seules. En effet, les dégradations AV incluant la dégradation V-DEB ont été perçues comme dégradant plus fortement la qualité que celles incluant V-GEL. Cette observation reflète bien les résultats obtenus lorsque les dégradations vidéo étaient présentées seules.

Les descripteurs *Couleur*, *Luminosité*, *Détail* ainsi que les descripteurs de haut-niveau n'ont pas permis d'apporter des éléments d'explications aux effets observées. Il semblerait que les influences potentielles de ces descripteurs aient été masquées par celles des autres critères.

7.5. CONCLUSIONS B2

Globalement, les résultats de l'expérimentation B2 ont montré que la **dégradation vidéo par réduction du débit (V-DEB) a été jugée comme la dégradation de qualité la plus forte** (diminuant le plus la note de qualité AV globale). Cela peut être constaté autant pour une présentation isolée de cette dégradation que pour une présentation combinée avec une dégradation audio (A-DEB ou A-PP). Concernant les dégradations audio, **la dégradation par perte de paquets (A-PP) a été jugée comme altérant plus fortement la qualité** que la dégradation par réduction du débit (A-DEB). Cet effet est également retrouvé lors d'une présentation combinée avec une dégradation vidéo (V-DEB ou V-GEL). Ce constat peut s'expliquer soit par des différences objectives entre les dégradations audio et vidéo (quantité de dégradations) soit par l'existence de régularités dans la perception de certaines dégradations. Les **dégradations liées au débit vidéo** (perte de résolution spatiale) **seraient perçues comme dégradant plus fortement la qualité** qu'une dégradation par rupture de la continuité du signal (saccades), indépendamment du type de contenu. Pour l'audio, à l'inverse, la rupture de la continuité (perte d'information) serait perçue comme plus gênante que les dégradations liées au débit (distorsion). Il se peut que la rupture de continuité du flux audio entraîne une diminution de l'intelligibilité en raison de la perte de certaines informations notamment verbales (phonème, syllabe, *etc*.), pour la vidéo, la perte de netteté constituerait une perte d'information visuelle plus importante que la rupture de continuité. Ce résultat peut également être mis en regard des spécificités des systèmes auditifs et visuels : l'acuité temporelle de l'audition et donc sa sensibilité (à la rupture de la continuité) est plus élevée que celle de la vision tandis que l'acuité spatiale de la vision et donc sa sensibilité (à la perte de résolution) est plus élevée que celle de l'audition.

Les résultats obtenus ont également permis d'observer **qu'une dégradation audio diminue davantage le niveau de qualité perçue lorsque la nature de la modalité**

dominante de la séquence était audio. Le même constat peut être apporté pour une dégradation vidéo et une modalité dominante vidéo. Par ailleurs, lorsque la modalité dominante est vidéo, la présence de saccades sur le signal vidéo (V-GEL) tend à dégrader plus fortement des séquences de dynamique forte (dégradation de l'information pertinente) tandis que la perte de résolution vidéo (V-DEB) dégraderait plus fortement des séquences de dynamique faible (dégradation des rares informations visuelles disponibles).

Enfin, la dégradation ***Désynchronisation*** **a été perçue comme dégradant la qualité lorsque celle-ci survenait sur des séquences présentant un contenu verbal en lien avec l'action affichée à l'écran (son diégétique)**. Lorsque ces deux conditions sont remplies, le phénomène de fusion des informations auditive et visuelle est alors altéré. Dans le cas contraire, les notes obtenues étaient proches de celles attribuées aux séquences présentées sans dégradations. Le cas échéant, la note de qualité ne permet pas de savoir si la désynchronisation était perçue ou non. Cette dernière observation semble confirmer l'importance d'ajouter une question spécifique à la perception de la désynchronisation lorsque cette dernière est étudiée.

L'expérimentation B2 a permis de mieux définir la relation entre contenu, caractérisé à partir d'un ensemble de descripteurs, et perception de qualité. Les influences du contenu observées, par exemple sur la perception de désynchronisation, sont des informations importantes à considérer lors de la sélection des séquences de test.

7.6. CONCLUSION GENERALE ET PERSPECTIVES

Les expérimentations B1 et B2 présentées dans ce chapitre ont permis d'étudier l'influence du contenu, d'une part, sur la *qualité d'expérience* du spectateur (en matière d'intérêt, de plaisir, *etc.*, B1) et d'autre part, sur la perception de qualité (B2).

L'expérience B1 a permis de disposer d'un corpus de contenus et de séquences de test audiovisuels caractérisés selon des descripteurs sémantiques (de bas et haut-niveau), techniques et hédoniques. B1 a montré que la caractérisation réalisée par l'expert (descripteurs de bas-niveau seulement) peut être considérée comme pertinente pour une description des contenus compréhensible par des participants naïfs en-dehors du descripteur de température de couleur. D'un point de vue méthodologique, cela autorise à ne mobiliser qu'un expert seul pour l'annotation des séquences ou contenus de test.

L'expérience B2 a permis de mieux comprendre les interactions entre le contenu, défini par les descripteurs proposés, et la perception de qualité. La modalité dominante, la dynamique, la relation audiovisuelle (diégétique) et l'expression sonore influencent notamment la qualité audiovisuelle perçue par le spectateur. Ce constat souligne l'importance de l'étape de sélection des séquences de test ainsi que la description préalable à réaliser. Par exemple, pour être étudiée, la désynchronisation devra être appliquée à des séquences présentant un segment audio à la fois verbal et diégétique.

Une seconde campagne de test pourrait être conduite afin de tester le questionnaire utilisé dans l'expérimentation B1 pour un protocole similaire à l'expérimentation B2. Cela permettrait d'étudier plus précisément les effets des dégradations non plus sur la perception de qualité seule mais également sur la *qualité d'expérience* finale du spectateur. Il pourrait également être intéressant d'élargir le cadre de la norme UIT-T P.911 aux évaluations des qualités audio et vidéo pour permettre l'étude du poids de chaque qualité à la perception de qualité audiovisuelle. De plus, dans le cadre du corpus de séquences étudiées, les descripteurs *Modalité* et *Dynamique* étaient fortement liés (dynamique faible associée à une dominance audio). Ce constat va dans le sens de celui de Hands (2004), cependant, cette association pourrait être l'expression d'une limite liée au questionnaire où le terme de dynamique réfère uniquement à la modalité vidéo. Une étude complémentaire où il serait demandé aux participants d'évaluer à la fois la dynamique audio et vidéo pourrait être envisagée pour approfondir ce point.

CHAPITRE VIII – EXPERIMENTATION C : ETUDE FINALE

8.1. INTRODUCTION GENERALE

Les résultats de l'expérimentation A ont souligné la difficulté d'obtenir des réponses significatives des indicateurs physiologiques et oculaires en réaction aux conditions expérimentales testées. Les expériences B1 et B2 ont permis de décrire le corpus de contenus de test et de mieux comprendre l'influence du contenu tant sur la qualité perçue que sur la *qualité d'expérience* du spectateur (envisagée sous l'angle de l'intérêt, du plaisir, *etc.*). Ces études ont également apporté des éléments d'explications pour interpréter les variations des mesures psychophysiologiques observées lors de l'expérimentation A. Au-delà d'une influence liée au contenu, l'effet des dégradations sur les mesures psychophysiologiques aurait également pu être masqué ou atténué par un certain nombre de facteurs, comme présenté en conclusion de l'expérience A, tels que la répétition des contenus (agacement/irritation, ennui, désengagement), la durée et le seuil des dégradations appliquées, la question du maintien du focus attentionnel sur la tâche de visualisation, l'effet du changement d'activité (repos *vs*. tâche, complétion du questionnaire *vs*. visualisation), *etc.* Le protocole nécessite donc d'être remanié pour favoriser l'observation de l'influence de la qualité sur les différentes mesures recueillies.

L'expérience décrite dans ce chapitre propose un protocole amélioré, c'est-à-dire tenant compte des facteurs considérés comme potentiellement préjudiciables à l'expression physiologique et/ou comportementale des dégradations.

8.2. OBJECTIFS

L'objectif de ce chapitre est de tester une évolution du protocole utilisé dans l'expérimentation A tenant compte des effets observés ou supposés des facteurs : *passation*, *changement d'activité* (amorce, avertisseur*)*, *habituation, engagement*, *matériel* et *niveau* (durée et seuils de dégradations). Des adaptations relatives aux mesures subjectives sont également proposées pour enrichir le questionnaire proposé et obtenir un retour plus complet de la perception de qualité et de la *qualité d'expérience* du spectateur. Les effets observés et non souhaités dans l'expérimentation A ainsi que les adaptations associées et testées dans l'expérimentation C sont résumés dans le Tableau 8.1 ci-après.

Tableau 8.1. Adaptations du protocole proposées à l'issue de l'expérimentation A et testées dans l'expérimentation C selon le type de mesures (Mes.) : Subjectives (SUBJ.) ou Psychophysiologiques. Les lignes grisées correspondent aux travaux déjà réalisés.

Mes.	Effets	Adaptations proposées et testées
SUBJ.	Limite du *questionnaire*	Ajouter une question spécifique à la perception de la désynchronisation
	Effet du facteur *niveau*	Augmenter le seuil de la désynchronisation
	Effet *contenu*	Caractériser les contenus de test
PSYCHOPHYSIOLOGIQUES	Effet du facteur *passation*	Proposer un ordre aléatoire de présentation/ Présenter une seule et unique fois les contenus
	Effet du facteur *activité*	Ajouter un contenu *amorce* Ajouter avertisseur de début de contenu
	Effet du facteur *habituation*	Varier patterns d'introduction des dégradations
	Effet du facteur *engagement*	Favoriser l'engagement sur la tâche
	Effet du facteur *matériel*	Solution de synchronisation
	Effet du facteur *niveau*	Augmenter les durées et seuils des dégradations appliquées
	Effet du facteur *analyse*	Varier les approches
	Effet *contenu*	Caractériser les contenus de test

Comme indiqué dans la section 4.4.3 (chap. IV), le format 3D vidéo augmente la probabilité d'une présence de fatigue visuelle par rapport à un format 2D vidéo (Eui Chul *et al.*, 2010 ; Jae-Hwan *et al.*, 2012). Il est donc possible de supposer que le traitement de la 3D (pour plus de détails sur la vision stéréoscopique voir annexe 2-A) implique un engagement de ressources plus important qu'un format 2D. Ainsi, le cumul d'un format 3D et de dégradations pourrait favoriser l'expression d'un effort mental puis de fatigue tel que mesurés par les indicateurs psychophysiologiques. De ce fait, le format 3D est considéré ici comme un facteur d'augmentation potentiel d'un effet préjudiciable des dégradations sur les mesures physiologiques et oculaires. L'idée est de proposer le terrain le plus favorable possible à l'étude des réponses psychophysiologiques en réaction aux fluctuations de qualité.

8.3. PARTICIPANTS

Trente-trois participants (17 femmes, 16 hommes), entre 18 et 52 ans, non porteurs de lunettes pour faciliter l'enregistrement des mesures oculaires, ont participé à l'expérience. Ils étaient rémunérés pour leur participation.

8.4. MATERIEL

8.4.1. CONFIGURATION GENERALE

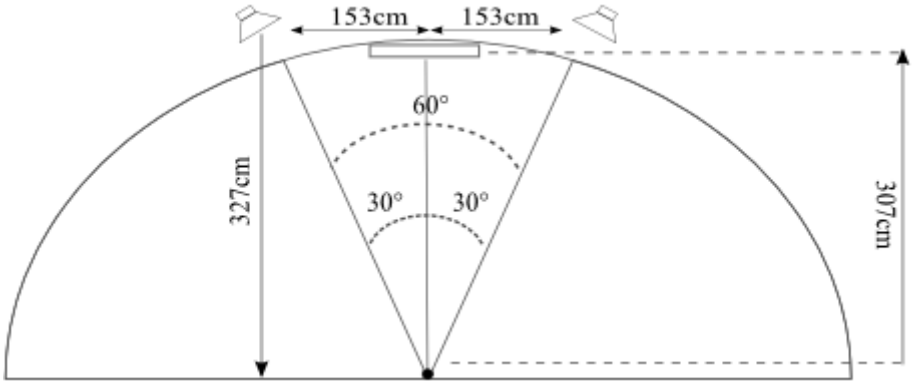


Fig. 8.1. Schéma de la configuration de la salle de test (193×376×505 cm) de l'expérimentation C. La place du participant est représentée par un point noir, l'écran est figuré par un rectangle.

Cette expérience a été conduite dans une pièce insonorisée. Les conditions de visualisation et d'écoute recommandées par la norme UIT-T P.911 (UIT, 1998) étaient respectées (pour plus de détails voir l'annexe 8-A). Le niveau de luminosité de la salle était inférieur à 20 lx. Les paramètres de brillance et de contraste de l'écran remplissaient les conditions de la norme UIT-R BT.814-2 (UIT, 2007). Un écran LCD 3D ready (200Hz, technologie passive c.-à-d. écran et lunettes polarisées) de marque LG (modèle 55LW650S) de 55" (140 cm) full HD a été utilisé pour afficher les contenus de test en format 3D full HD (1920×1080p, 60 Hz). Aucune norme ne spécifie actuellement la distance de visualisation pour des tests impliquant une visualisation 3D. Ainsi, la distance de visualisation, en accord avec la norme UIT-T P.911, a été fixée à 4,5 fois la hauteur de l'écran (307 cm).

Les haut-parleurs utilisés dans cette expérience étaient de marque Genelec et de modèle 8040A. Les deux haut-parleurs étaient réglés à une hauteur de 94,5 cm avec un niveau d'écoute, mesuré à la tête du participant pour simuler les conditions réelles d'écoute, d'environ 80 dBA comme indiqué dans la norme UIT-T P.911. Le réglage du volume sonore a été effectué à partir des fichiers AV de test normalisés. Le son ambiant de la salle de test était inférieur à 30 dB. La Figure 8.1 ci-dessus présente la configuration de la salle de test qui respectait les recommandations de la norme UIT-R BS.1286 (UIT, 1997).

8.4.2. CONFIGURATION TECHNIQUE

La configuration et le matériel (ordinateur capable de lire des formats full HD) de diffusion des signaux audio et vidéo étaient identiques à ceux de l'expérimentation B1 (voir § 7.2.4.2, chap. VII). Les séquences audiovisuelles ont également été diffusées *via* le *player* multimédia Windows Media (WM). Son utilisation présente trois avantages majeurs : création facile de *playlist* (modifiables directement en format *.txt* pour une présentation aléatoire des contenus), le déclenchement à distance sans délai (*via* réseau local) et par conséquent, la possibilité de

synchroniser les logiciels entre eux (logiciels de lecture des contenus audiovisuels et d'enregistrement des différentes mesures psychophysiologiques).

8.4.3. SOLUTION DE SYNCHRONISATION

La Figure 8.2 ci-après présente la solution de synchronisation entre les différents logiciels utilisés. Afin de garantir la précision des fenêtres temporelles comprenant les mesures extraites des fichiers de données brutes et devant correspondre à chaque période dégradée et non dégradée, il était nécessaire de synchroniser le déclenchement de l'ensemble des logiciels.

La lecture des contenus nécessitant des ressources assez importantes, un ordinateur spécifique était dédié à cette tâche (PC contenu). En parallèle, un ordinateur était alloué à la gestion des mesures (PC gestion) et un ordinateur à l'enregistrement des mesures oculaires (PC faceLAB). Des solutions techniques et logicielles ont ensuite dû être mises en place pour assurer la synchronisation entre la diffusion du contenu et les mesures enregistrées. La synchronisation était assurée grâce à un module indépendant de déclenchement appelé *Télécommande*. Ce module permettait de déclencher la lecture des contenus AV (PC contenu) à partir du PC gestion *via* un réseau local.

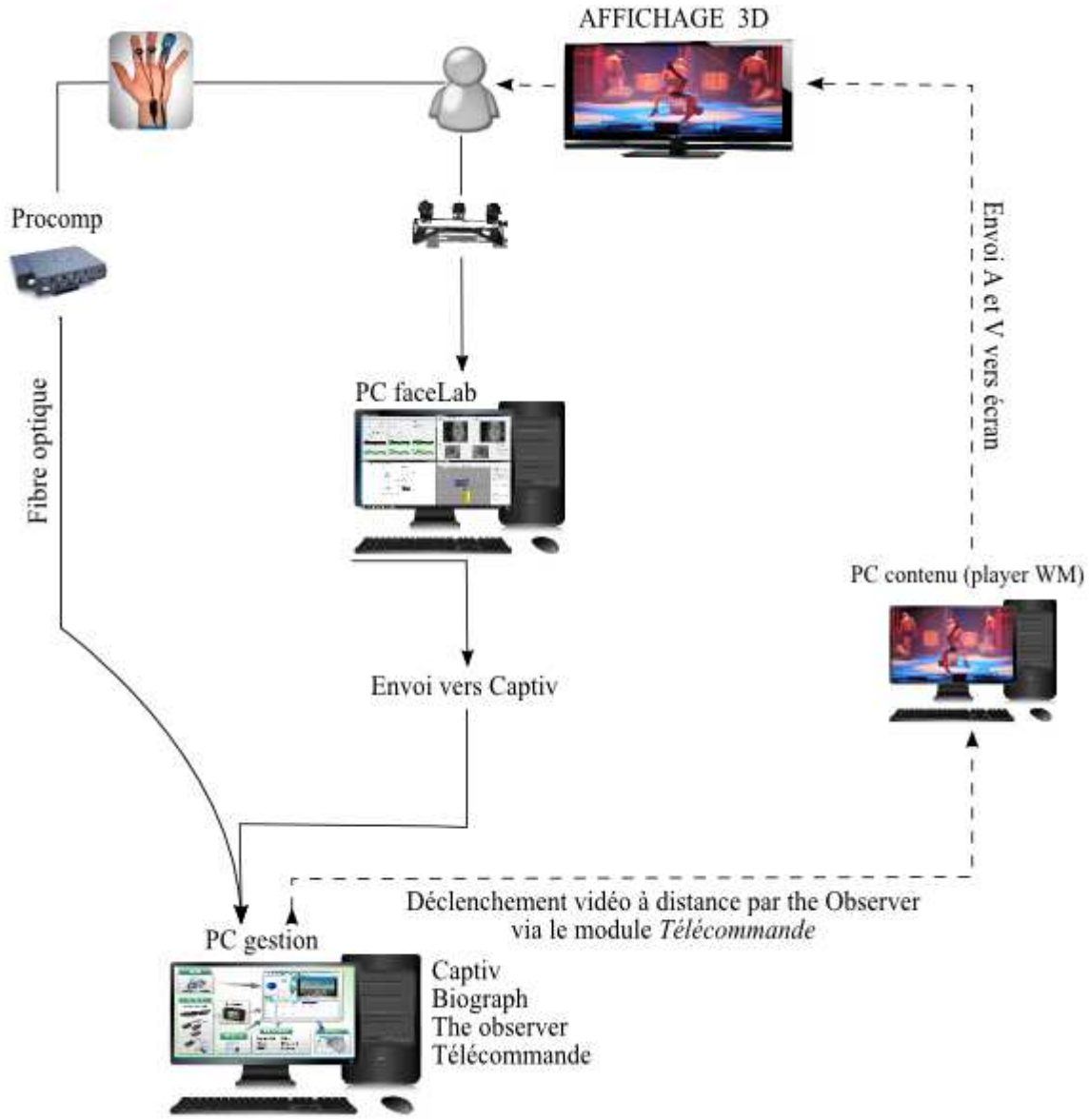


Fig. 8.2. Solution automatisée de synchronisation : déclenchement synchrone des différents logiciels *via* le logiciel The Observer. FaceLAB et WM étaient déclenchés par l'intermédiaire d'un second module (Télécommande). Celui-ci enregistrait également l'heure du déclenchement de chaque logiciel. Les logiciels Captiv[21], The Observer et Biograph étaient installés sur un même ordinateur (PC gestion) tandis que WM et faceLAB se trouvaient sur des ordinateurs dédiés à leurs fonctionnements respectifs.

Concrètement, le déclenchement simultané des différents logiciels a été réalisé au moyen du logiciel The Observer®XT (Noldus Information Technology). Celui-ci permettait de

[21] Le logiciel d'acquisition de mesures Captiv 7000 (TEA ergo) devait permettre l'acquisition simultanée des mesures physiologiques et oculaires au sein d'un même logiciel ainsi que l'insertion de marqueurs temporels signalant le début et la fin de chaque période (dégradée ou non dégradée). La version proposée de Captiv était une version non aboutie et certaines erreurs (enregistrements ou exports des mesures pas toujours possibles, impossibilité d'ajouter les mesures physiologiques, déclenchement des marqueurs temporels erronés, *etc.*) n'ont pas permis l'utilisation des données enregistrées.

déclencher à la fois le logiciel Biograph (installé sur le même ordinateur) et les logiciels faceLAB et WM (installés sur deux ordinateurs dédiés, ceux-ci ne pouvant supporter que le fonctionnement de leur propre logiciel). Pour ces deux derniers logiciels, le déclenchement était effectué *via* le module *Télécommande* (1 ms de délai), The Observer ne pouvant pas réaliser simultanément plusieurs déclenchements. Une illustration du module *Télécommande* est donnée par la Figure 8.3 ci-dessous. Un délai fixe de 300 ms entre le déclenchement de The Observer et celui des autres logiciels a dû être pris en compte par le module *Télécommande*. Une seconde fonction de ce module était de marquer l'heure de déclenchement de chaque logiciel (précision à la ms) sur l'ensemble des ordinateurs de l'expérimentation (PC contenu, PC gestion et PC faceLAB). Ces marqueurs temporels étaient enregistrés dans un fichier *.txt* sur chacun des postes. Cette information a permis de resynchroniser les mesures *a posteriori* et de réaliser avec précision la correspondance entre le déroulé de la passation (pattern de dégradations) et les mesures enregistrées.

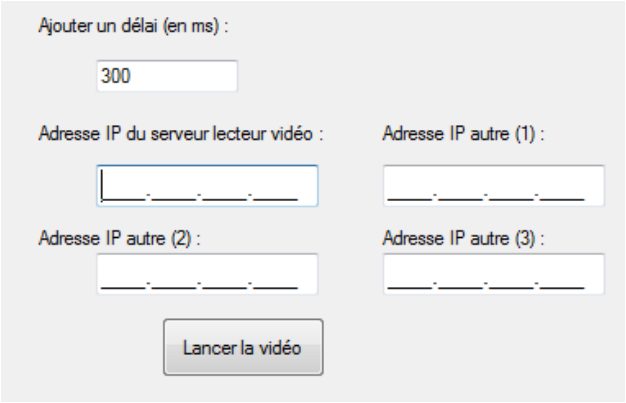


Fig. 8.3. Illustration du module *Télécommande* pour le déclenchement des logiciels WM et faceLAB ainsi que l'enregistrement de l'heure de déclenchement de chacun des logiciels.

8.4.4. RECUEIL DES DONNEES

Les données oculaires étaient capturées par deux optiques (eye tracker faceLAB5™) et acheminées *via firewire* vers un ordinateur dédié[22]. La capture des signaux physiologiques était identique à l'expérimentation A (logiciel *Biograph Infiniti* installé sur un second ordinateur[23]). Cependant, contrairement à la précédente expérimentation, le capteur de conductivité électrodermale fourni avec l'outil de mesure (SC-Flex/Pro, électrodes Ag-AgCl) n'a pas été utilisé. Dans cette expérience, deux électrodes (Ag-AgCl, modèle V91-01, 8mm, Coulbourn Instruments) reconnues pour leur fiabilité ont été préférées. Ces électrodes nécessitaient l'utilisation d'un gel isotonique (Gel 101, Biopac Systems, Inc.). Comme recommandé par Fowles *et al.* (1981, sect. 3.3.4, chap. III), des disques adhésifs double-face ont été utilisés pour optimiser le maintien des électrodes et la surface de contact avec la peau.

[22] Dell, Intel Core i7, 2,80 Ghz, ram :4 Gb, OS: Windows XP (32 bits)

[23] Dell ,Intel Xeon W3540, 2,93GHz, ram : 12 Gb, OS windows XP (64 bits)

L'ensemble des signaux physiologiques étant influencé par la température extérieure (Venables et Christie, 1980) et celle-ci ne pouvant être contrôlée (pas de thermostat ou de climatiseur présents dans la salle de test), la température a été mesurée et enregistrée tout au long de la campagne de test au moyen d'un système d'enregistrement sans fil (Arexx TL-500, ±0,5°C).

L'ensemble des ordinateurs utilisés était placé dans une salle de régie annexée à la salle de test. Pour éviter toute interférence avec les mesures physiologiques, les téléphones portables étaient interdits dans la salle de test.

8.5. STIMULI

Le corpus de contenus audiovisuels utilisé dans cette expérience était identique au corpus caractérisé par l'expert (chap. VII) à savoir :

- **Danse :** extrait du ballet *Balé de Rua* (14 min 21),
- **Documentaire** : documentaire entier sur Jean-Marc Mormeck (12 min 25),
- **Opéra** : extrait d'une adaptation de *Don Giovanni* (12 min 36),
- **Sport** *:* extrait de la finale de Rolland Garros 2011 (12 min),
- **Théâtre** : extrait d'une adaptation des *Fourberies de Scapin* (10 min 29).

Un aperçu des contenus AV de test est présenté par la Figure 8.4 ci-dessous.

Fig. 8.4. Aperçu des contenus de test avec de gauche à droite : Documentaire, Opéra, Sport, Danse et Théâtre.

Les cinq contenus de test étaient présentés au format 3D vidéo (présentation Side-by-Side) et stéréo (audio) full HD 1080p. Les contenus ont préalablement été encodés à 15 Mbps (AVC-x264, .avi, compression simultanée de l'image gauche et droite pour une compression symétrique). Le format 3D était natif (c.-à-d. tourné à l'aide de caméras 3D) pour les cinq contenus ; il n'était donc pas reconstruit *a posteriori* (tourné en 2D puis converti en 3D), L'encodage des contenus a été nécessaire pour permettre à l'ordinateur la lecture de séquences audiovisuelles aussi longues et la restitution au format full HD. Le seuil d'encodage a été choisi pour se situer au-dessus des normes de diffusion actuelle de TV HD

(actuellement de 6 Mbps[24], sect. 1.2, chap.1). Rappelons que la présentation 3D a été choisie pour sa capacité à générer de la fatigue visuelle, l'utilisation d'un tel format pourrait alors proposer un contexte plus favorable pour observer des réponses psychophysiologiques en réaction aux fluctuations de qualité.

Quatre dégradations, choisies sur la base des résultats obtenus à l'expérimentation B2 (chap. VII), ont été appliquées dans cette étude. Les dégradations audio et vidéo sélectionnées correspondaient à celles ayant reçu les notes de QAV les plus basses (augmentation des niveaux de dégradations par rapport à l'expérimentation A). Il s'agissait des dégradations :

- **V-DEB** (vidéo) : réduction du débit vidéo variable, entre 94 et 550 Kbps, selon le contenu,
- **A-PP** (audio) : 10% de perte de paquets,
- **V-DEB * A-PP** (audiovisuelle) : combinaison des dégradations vidéo et audio,
- **D** (audiovisuelle) : désynchronisation image/son avec 1500 ms de retard de l'image sur le son.

Chaque dégradation était appliquée durant une minute (augmentation de la durée de dégradation par rapport à l'expérimentation A), chaque contenu comportait donc un total de quatre périodes dégradées. La Figure 8.5 présente le pattern d'introduction des dégradations pour chaque contenu.

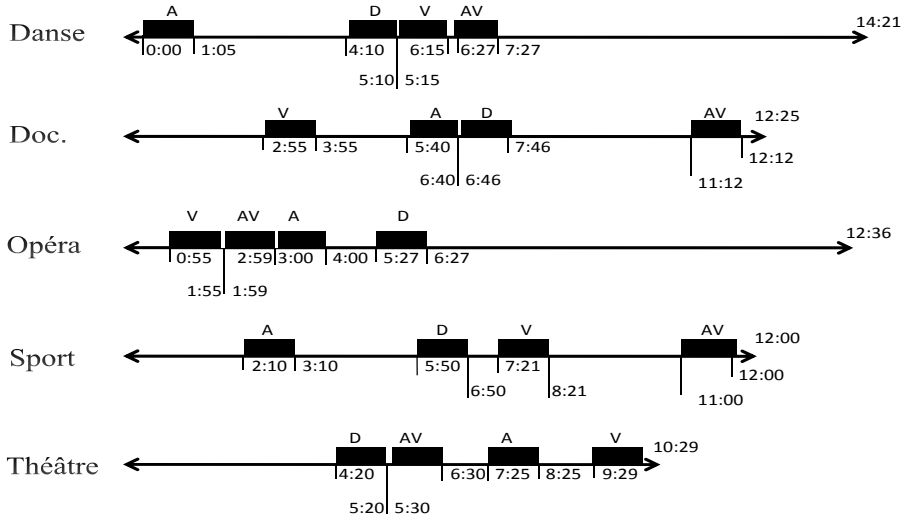


Fig. 8.5. Patterns d'introduction des dégradations pour chaque contenu. D représente la désynchronisation, A la dégradation audio, V la dégradation vidéo et AV la dégradation audiovisuelle.

[24] Actuellement, le débit de diffusion d'un flux 3D est fixé selon la libre appréciation du diffuseur. A titre indicatif, le débit choisi par le réseau Orange est le même que celui de la diffusion HD (pour des contraintes techniques) fixé à environ 6 Mbps pour l'ADSL et 11 Mbps pour la TNT.

Chaque dégradation était présentée une seule fois au sein de chaque contenu. Au total, un contenu donné était divisé en huit (Danse, Sport et Théâtre) ou neuf périodes (Documentaire et Opéra), les périodes non dégradées étaient de durées variables (quelques secondes à plusieurs minutes). L'irrégularité des patterns, différents pour chaque contenu, devait permettre d'éviter la prévisibilité d'apparition des dégradations.

A chaque contenu correspondait un pattern de dégradations qui lui était propre, le pattern d'introduction des dégradations d'un contenu donné étant identique pour l'ensemble des participants. En revanche, l'ordre de présentation des contenus était aléatoire pour chaque participant.

Le pattern d'introduction des dégradations a été réalisé sur la base de résultats de l'expérimentation B2. Chaque période dégradée (1 min) englobait la séquence de dix secondes ayant obtenu, pour une dégradation donnée, la note MOSAV la plus basse lors de l'expérience B2 (augmentation des seuils et de la durée des dégradations). Cela devait permettre de disposer de manière certaine d'au moins dix secondes pour lesquelles les dégradations étaient consciemment perçues.

8.6. OBSERVABLES

8.6.1. MESURES SUBJECTIVES

L'ensemble du questionnaire de l'expérimentation B1 permettant l'évaluation des catégories Hédonique, Sémantique et Technique a été soumis aux participants à l'exception de descripteur de *Température de couleur* peu adapté à un public non expert en technique audiovisuelle. De plus, une série de quatre questions (catégorie Perception) a été ajoutée afin d'obtenir plus de précisions sur la perception des dégradations A, V, AV ou D et du format 3D. Ces questions étaient relatives à la détection de la dégradation (A, V, AV, D) ou d'une gêne associée à la 3D (gêne 3D) et au niveau de certitude associé à cette détection/gêne. Si et seulement si, une dégradation ou une gêne 3D était détectée, alors les participants devaient répondre à deux questions supplémentaires portant sur l'impact de la dégradation/gêne 3D sur la compréhension du contenu et sur le sentiment d'émotions négatives (agacement, énervement, stress, frustration, *etc.*). Au total, les participants devaient répondre à un ensemble de trente et une questions. Le questionnaire est présenté dans l'annexe 8-B. L'ensemble des critères subjectifs évalués dans cette étude est présenté dans le Tableau 8-2 ci-après.

Tableau 8.2. Récapitulatif de l'ensemble des observables de l'expérience subjective des spectateurs pour l'expérimentation C.

Observable	Echelle	Catégorie
QAV, QV, QA	ACR 9 points - 5 items : excellent-bon-satisfaisant-médiocre-mauvais	Qualité
Intérêt	faible-modéré-fort	Hédonique
Valence	9 niveaux (SAM)	Hédonique
Arousal	9 niveaux (SAM)	Hédonique
Quantité d'information	faible-modérée-forte	Sémantique
Compréhension	faible-modérée-forte	Sémantique
Modalité	A, V, AV	Sémantique
Dynamique contenu	faible-modérée-forte	Sémantique
Luminosité	faible-modérée-forte	Technique
Détection (D, A, V, AV, 3D)	binaire (UIT P.805) : oui - non	Perception
Certitude détection (D, A, V, AV, 3D)	faible-modérée-forte	Perception
Gêne Compréhension (D, A, V, AV, 3D)	5 niveaux - 5 items : pas du tout-légèrement-moyennement-beaucoup-extrêmement	Perception
Emotions négatives (D, A, V, AV, 3D)	5 niveaux - 5 items : pas du tout-légèrement-moyennement-beaucoup-extrêmement	Perception

Les observables sont classés selon les catégories Qualité, Technique, Sémantique, Hédonique et Perception. Les échelles binaires et en 5 points (« Pas du tout » à « Extrêmement ») sont issues de la norme UIT-T P.805 (2007) relative aux méthodes subjectives d'évaluation de qualité en contexte conversationnel. L'évaluation de la gêne liée au format 3D est notée 3D, la désynchronisation est notée D.

8.6.2. MESURES PHYSIOLOGIQUES ET OCULAIRES

Cinq indicateurs du comportement oculaire ont été mesurés dans le cadre de cette expérimentation : PERCLOS, Diamètre Pupillaire (DP), fréquence et durée de fermeture de l'œil (Eye Blink : EBfreq et EBdur) et le nombre de Saccades (SAC). Ce dernier indicateur a été ajouté pour apporter un indicateur supplémentaire de fatigue. Pour les indices physiologiques, la conductance cutanée (AED), la volumétrie sanguine périphérique (VSP), la fréquence cardiaque (FC) et la température cutanée périphérique (TCP) ont été mesurées.

Les fréquences d'échantillonnage était de 32 Hz pour les données d'AED, de TCP, de VSP et de FC (calculée, *a posteriori*, à partir du VSP) et de 60 Hz pour l'ensemble des indices oculaires. Les observables physiologiques et oculaires et les outils de recueil utilisés sont synthétisés par le Tableau 8.3 ci-après.

Tableau 8.3 : Synthèse des différents observables et outils de recueil pour chaque type de mesures étudié.

Type de mesure	Observable	Recueil
PHYSIOLOGIQUE	AED	Paire d'électrodes
	FC	Pléthysmographe
	VSP	
	TCP	Thermistor
OCULAIRE	DP	Eye tracker
	EBdur	
	EBfreq	
	PERCLOS	
	SAC	

8.7. PROTOCOLE

Dès leur arrivée, les capteurs pour la mesure de l'AED (électrodes placées sur les phalanges médiales de l'index et du majeur), de TCP et de VSP étaient installés sur la main non dominante des participants (voir chap. III). Comme indiqué dans l'expérience A, cette installation précoce avait pour objectif d'une part, de stabiliser les mesures physiologiques et d'autre part, d'habituer le participant au port des capteurs.

Après l'installation des capteurs, une présentation orale du déroulement de l'expérience était donnée. Le contexte général d'évaluation présenté aux participants était celui de l'évaluation d'un service de diffusion de contenus 3DTV encore à l'essai (aucun contexte d'évaluation n'était précisé dans les précédentes expériences, l'objectif était ici d'obtenir une plus grande implication des participants dans la tâche d'évaluation). Les participants pensaient donc devoir juger d'un système de diffusion en cours d'élaboration, l'état non abouti de ce service devait permettre de justifier les variations de qualité prévues dans le protocole.

Les participants devaient ensuite évaluer leur niveau de fatigue sur une échelle graduée à sept niveaux et trois items (voir annexe 8-C pour plus de détails). Après cette évaluation, les synopsis (annexe 8-D) décrivant brièvement les cinq contenus de test étaient remis aux participants. Cette étape avait pour objectif d'engager plus largement le participant dans l'activité de visualisation de contenu en précisant le contexte narratif général des extraits présentés (par exemple, le bref résumé des Fourberies de Scapin devait permettre au participant de ne pas être « perdu » lors de la présentation de l'extrait et d'éviter un désengagement de la tâche de visualisation). Les contenus devaient ensuite être classés selon leur ordre de préférence (annexe 8-E). Ces informations étaient à nouveau recueillies après la passation du test afin d'obtenir un retour sur les modifications éventuelles du niveau de fatigue ou de préférence avant et après le test. Par exemple, il est envisageable que l'ordre de préférence des contenus ait été réévalué après la découverte des extraits. La disponibilité de cette information fournit une première indication du niveau d'attractivité d'un contenu pouvant influencer les mesures recueillies.

Enfin, les consignes et le questionnaire papier étaient remis et présentés au participant (visibles respectivement dans les annexes 8-F et 8-B). Après cette étape, le calibrage et la création du modèle de tête pour l'enregistrement des mesures oculaires étaient réalisés suivant la procédure décrite dans la section 3.6.1 (chap. III). Le port des lunettes 3D (passive) ne constituait pas une gêne pour le recueil de ces mesures (participants non porteurs de lunettes correctives). Cette étape durait en moyenne quinze à vingt-cinq minutes. Pour améliorer le confort des participants et favoriser un maintien naturel de la position de tête, une chaise ergonomique (dossier haut, roulettes fixes) a été utilisée dans cette expérience. Les mesures étant particulièrement sensibles au mouvement, les consignes demandaient également aux participants d'éviter, dans la mesure du possible, les mouvements du bras ou de la main supportant les capteurs. Toutefois, les participants avaient la possibilité de repositionner ou de se dégourdir la main ou le bras durant les phases de complétion de questionnaire. Un coussin permettait d'améliorer le confort et le maintien de la position de la main tout au long du test.

A l'issu de cette première phase de prise de connaissance du contexte et de préparation, suivait une phase de relaxation pendant laquelle le participant avait pour consigne de se détendre (aucune activité n'était demandée). Cette phase de repos durait cinq minutes (selon Gerin *et al.*, 1994, un enregistrement de cinq minutes serait suffisant pour obtenir une mesure stable, voir sect. 3.5, chap. III) pendant lesquelles l'activité physiologique était enregistrée (baseline). La phase de préparation et de présentation du test (20 à 30 min) et l'enregistrement de la baseline devait permettre de laisser s'écouler un laps de temps suffisant pour l'élimination d'éventuels résidus d'activités pré-test potentiellement néfastes (biais dans les mesures) tels que la prise de café, de cigarettes ou la réalisation d'un effort avant le test (même si proscrites deux heures avant le test).

Avant la présentation des cinq contenus audiovisuels de test, une amorce audiovisuelle 3D d'une minute était présentée au participant. L'objectif de cette amorce était de préparer le participant à l'activité de visualisation. Ce contenu présentait un cube en mouvement (avec variation de couleurs) sur fond gris et accompagné de bips sonores (normalisés autour de 80 dBA). Cette étape devait permettre d'amortir l'effet potentiel du changement d'activité de la phase de repos à celle de visualisation constatée sur l'AED et le VSP lors de l'expérimentation A.

Enfin, la dernière phase correspondait à la passation du test, c'est-à-dire la visualisation et l'écoute des contenus AV/3D de test. Ceux-ci étaient présentés de manière aléatoire pour chaque participant. Entre chaque contenu, les participants disposaient d'une pause de cinq minutes (signalée par un écran noir) pour compléter le questionnaire subjectif. Dans le cas où le questionnaire était terminé avant la fin du temps imparti, les participants étaient invités à se relaxer ainsi qu'à se repositionner en face des optiques de l'*eye tracker*. Cela permettait de contrôler la position et la qualité du *tracking* avant la reprise de l'activité de visualisation. Le participant était averti du commencement de chaque nouveau contenu par un avertisseur visuel et sonore, une minute puis six secondes (décompte cinématographique 5-4-3-2-1-0) avant son début effectif. L'amorce était également annoncée par la présence du

décompte. Ces « avertisseurs » devaient remplir deux objectifs. Premièrement, il s'agissait de minimiser une situation potentiellement stressante liée à la gestion du temps durant la phase de complétion du questionnaire, les participants étaient prévenus avant le début du test de la présence de ces avertisseurs. Deuxièmement, les avertisseurs devaient permettre aux participants d'anticiper l'activité de visualisation (placement du regard sur l'écran) et ainsi, de diminuer l'effet d'un changement trop brutal d'activité.

La durée totale du test était comprise entre 1 h 45 min et 2 h, dont 1 h 27 min de phase de visualisation et complétion de questionnaires. La Figure 8.6 présente le déroulement et le chronogramme de l'expérimentation C.

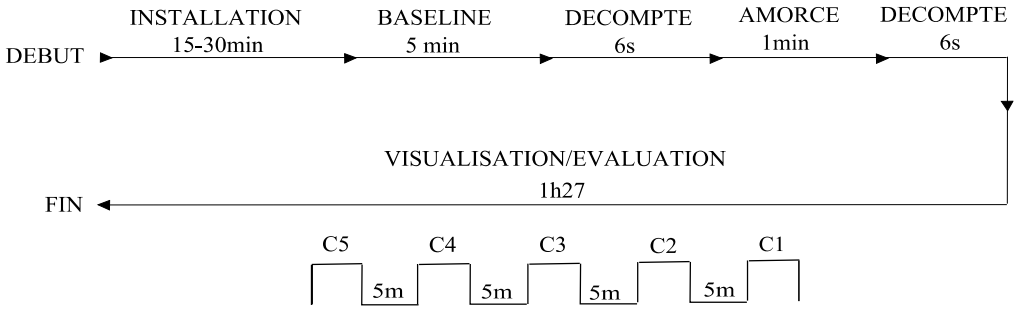


Fig. 8.6. Déroulement et chronogramme de l'expérimentation C : de l'installation des capteurs à la visualisation des contenus (C). Une pause de 5 min permettait aux participants de compléter le questionnaire.

Le protocole mis en place dans cette expérience tenait compte de l'ensemble des adaptations proposées à l'issu de l'expérimentation A.

8.8. HYPOTHESES

Le questionnaire proposé devrait permettre de compléter les notes MOS récoltées. Les hypothèses testées, concernant les mesures psychophysiologiques, dans cette expérimentation étaient identiques à celles de l'expérimentation A, à savoir :

- **H0p** : la présence de dégradations audio et/ou vidéo (désynchronisation, réduction du débit) pourrait être à l'origine d'un effort mental supplémentaire (pour décoder et interpréter le message dégradé) puis d'un état de fatigue, visibles à travers des modifications des patterns physiologiques et/ou oculaires,
- **H1p** : un effort mental supplémentaire lié à la présence de dégradations audio et/ou vidéo conduirait à une activation majoritaire du système nerveux sympathique traduite par une augmentation des indices AED, FC, DP et une diminution des indices VSP, TCP, EBfreq et EBdur,
- **H2p** : un état de fatigue consécutif à l'effort mental lié à la présence de dégradations audio et/ou vidéo serait traduit par une augmentation du PERCLOS, de EBfreq et de EBdur et une diminution de SAC.

8.9. RESULTATS

8.9.1. PREPARATION DES DONNEES

Au total, trente-trois participants ont pris part à l'expérimentation C. Chaque jeu de données, subjectif, physiologique ou oculaire a été considéré et traité séparément.

Sept participants ont été rejetés pour l'analyse du jeu de données physiologiques en raison de la présence de valeurs aberrantes ou d'artefacts trop nombreux (mouvements, mauvais maintien des électrodes –mains moites-, *etc.*). Au total, l'analyse statistique portant sur les mesures physiologiques a été réalisée à partir des données de vingt-six participants.

Pour le jeu de données oculaires, quatorze participants ont été exclus du jeu de données initial en raison de difficultés d'enregistrement (difficulté de détection de la pupille, artefacts de mouvement, *etc.*), portant à dix-neuf le nombre de participants dont les mesures ont été considérées comme valides. Le critère de rejet des participants a été décidé selon l'indicateur du niveau de qualité de la mesure (niveau de fiabilité) fourni par faceLAB. En effet, chacun des points de mesure (60 mesures/s) bénéficie d'un indice de qualité pouvant varier de zéro (aucune donnée) à trois (qualité de la mesure optimale). A partir de la moyenne de cet indice, obtenue pour chaque période étudiée, les données recueillies pour un participant étaient incluses dans le jeu de données si et seulement si plus de 70% des moyennes le concernant étaient supérieures à l'indice 2.

Aucun participant n'a été rejeté pour l'analyse du jeu de données subjectives. Le nombre de participants dont les mesures ont été retenues pour chaque jeu de données est récapitulé par le Tableau 8.4.

Tableau 8.4. Nombre de participants dont les mesures ont été retenues pour l'analyse statistique à partir des données subjectives, physiologiques ou oculaires.

TYPE DE MESURES	PARTICPANTS RETENUS
SUBJECTIVES	33
PHYSIOLOGIQUES	26
OCULAIRES	19

Les figures ci-dessous présenteront un intervalle de confiance à 95%. Les analyses *post-hoc* ont été effectuées à l'aide de tests HSD de Tukey.

8.9.2. MESURES SUBJECTIVES

8.9.2.1. FATIGUE

Un test de *Student* a mis en avant une augmentation significative du niveau moyen de fatigue entre les évaluations pré et post-test ($t(33) = 8{,}48$, $p < 0{,}05$). La Figure 8.7 illustre cette influence. L'effet de la classe d'âge (18-25, 25-35 et >35 ans) a également été vérifié au moyen d'une ANOVA à 1 facteur. Cette dernière n'a pas révélé d'effet significatif de la

classe d'âge sur les variables dépendantes « Niveaux de fatigue pré-test » (F (2, 31) = 0,61, p =0,55) et « Niveaux de fatigue post-test » (F (2, 31) = 1,23, p =0,31).

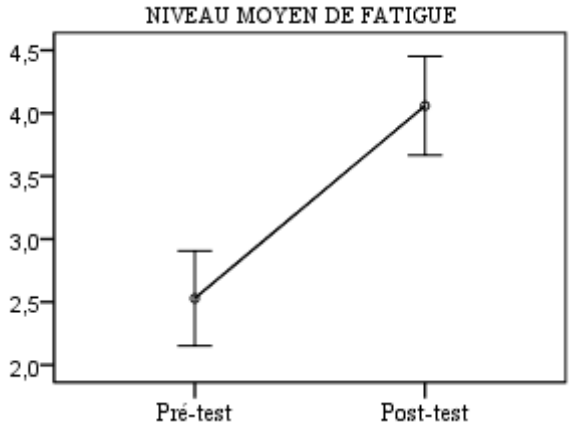


Fig. 8.7 : Evolution du niveau moyen de fatigue mesuré avant et après le test sur une échelle allant de 1 (« Pas du tout fatigué ») à 7 (« Extrêmement fatigué »).

8.9.2.2. PREFERENCES

La Figure 8.8 ci-après représente la position médiane de chaque contenu obtenue au classement de préférence avant et après la phase de visualisation. Pour plus de précision, la répartition des effectifs est apportée dans l'annexe 8-G.

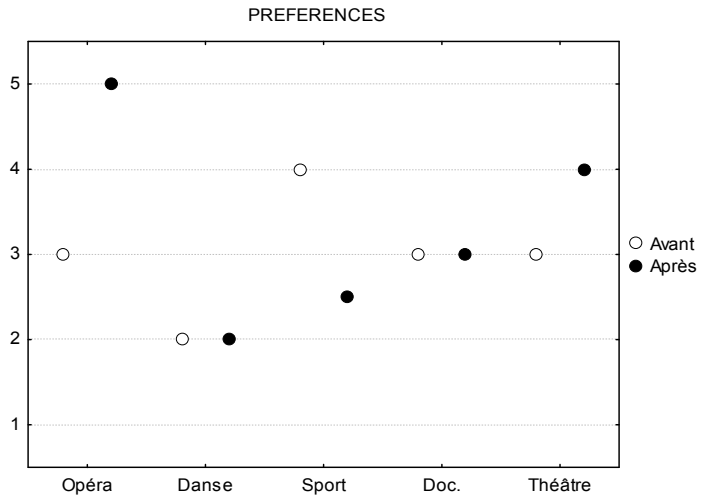


Fig. 8.8. Position (selon la médiane) de chaque contenu obtenue au classement de préférence, de 1 (contenu préféré) à 5 (contenu le moins préféré), avant et après la phase de visualisation.

Globalement les participants n'ont pas montré de préférence initiale marquée pour l'un ou l'autre des contenus. Ce constat permet de penser que les évaluations des critères ne devraient pas être influencées par des attentes spécifiques (considérées pour l'ensemble des participants) envers un contenu en particulier.

En revanche, les participants ont massivement (70% des participants) placé le contenu *Opéra* en dernière position (rejet) tandis que le contenu *Danse* a majoritairement été classé en

première position (adhésion) après la visualisation des contenus. Ce classement tend à confirmer les observations réalisées lors de l'étude B1, portant sur l'évaluation de courts extraits de ces mêmes contenus et pour lesquels les séquences du contenu *Opéra* avaient reçu les notes les plus basses d'intérêt, de valence et d'arousal. A l'inverse, les séquences du contenu *Danse* avaient obtenu les notes les plus hautes pour ces mêmes descripteurs.

8.9.2.3. CATEGORIE HEDONIQUE

La Figure 8.9 présente les niveaux moyens d'intérêt, de valence et d'arousal obtenus pour chaque contenu. Comme constaté lors de B1, *Opéra* a été à l'origine d'une expérience hédonique globale plutôt négative en matière d'intérêt, de valence et d'arousal. Une MANOVA considérant la variable indépendante « Contenu » et les variables dépendantes « Intérêt » à trois modalités (faible, modéré, fort), « Plaisir » et « Arousal » à neuf modalités (échelle de neuf niveaux) a indiqué un effet du contenu sur les descripteurs hédonique : F (12, 431,55) = 6,44, p <0,001.

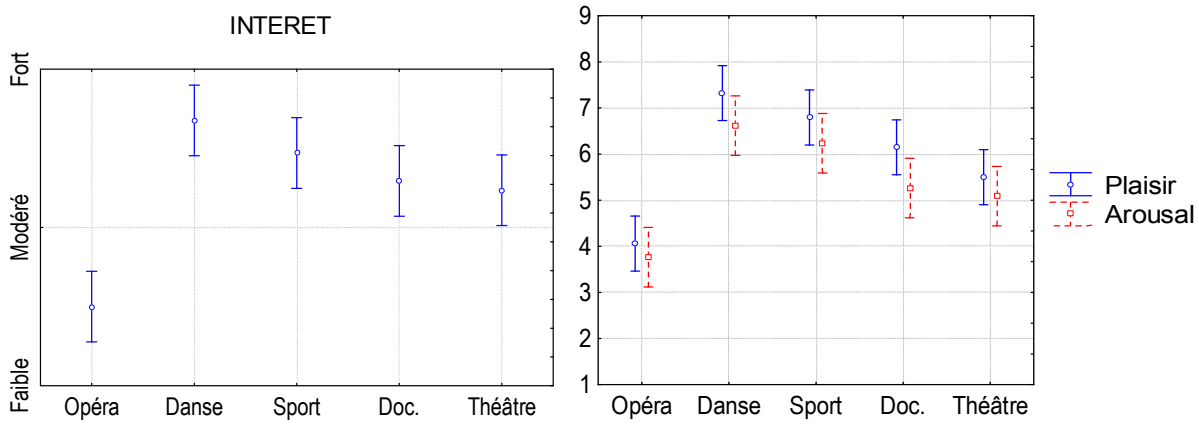


Fig. 8.9. Influence du contenu sur le niveau moyen obtenu pour les descripteurs Intérêt, Plaisir et Arousal de la catégorie Hédonique.

De manière générale, il semble que les descripteurs de la catégorie Hédonique reflètent le classement des contenus par ordre de préférence.

8.9.2.4. CATEGORIE SEMANTIQUE

La Figure 8.10 ci-dessous présente la répartition des effectifs pour l'évaluation du descripteur *Modalité* et les niveaux moyens obtenus pour les descripteurs *Quantité d'information*, *Compréhension* et *Dynamique de contenu* annotés après la visualisation de chaque contenu.

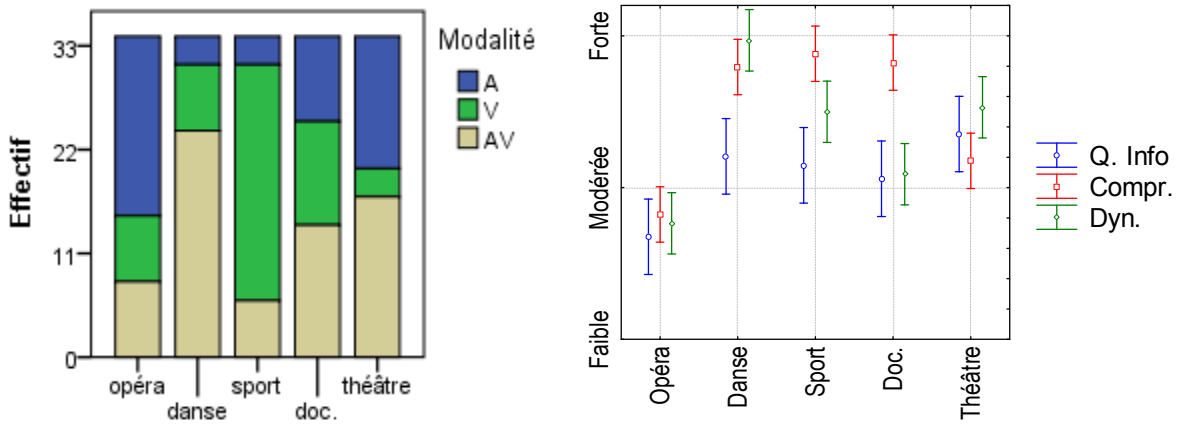


Fig. 8.10. Influence du contenu sur l'évaluation de la modalité dominante présentée selon la répartition des effectifs et le niveau moyen obtenu pour les descripteurs *Quantité d'information* (Q.info), *Compréhension* (Compr.) et *Dynamique* (Dyn.).

Concernant l'évaluation de la modalité dominante, le contenu *Danse* a majoritairement été annoté sans modalité dominante (modalité AV) tandis que les contenus *Opéra* et *Sport* ont respectivement été caractérisés par une modalité dominante audio et vidéo. La détermination de la modalité dominante pour les contenus *Documentaire* et *Théâtre* est moins évidente, toutefois, le premier peut être considéré comme AV avec une prédominance de la vidéo et le second comme AV avec une prédominance de l'audio. Ce résultat correspond globalement à la caractérisation experte de la modalité dominante (d'après le mode obtenu pour toutes les séquences analysées d'un contenu donné).

Le contenu a également influencé l'évaluation des autres descripteurs sémantiques comme l'indique la figure de droite. Cette observation a été confirmée par une série d'ANOVAs conduite selon la variable indépendante « Contenu » et la variable aléatoire « Participant » sur les variables dépendantes « Quantité d'information», « Compréhension» et « Dynamique » à trois modalités (faible, modérée, forte) avec respectivement $F\,(4, 132) = 4{,}18$, $p < 0{,}01$, $F\,(4, 132) = 27{,}57$, $p < 0{,}001$ et $F\,(4, 132) = 21{,}67$, $p < 0{,}001$. Les résultats n'ont pas révélé d'effet significatif de la variable « Participant ». L'observation de la figure permet plusieurs constats. Globalement, les descripteurs sémantiques représentés ne sont pas des notions corrélées, les participants ont été en mesure d'évaluer distinctement les critères de compréhension, de quantité d'information et de dynamique. Le niveau de compréhension est plutôt bon (de modéré à fort) quel que soit le niveau perçu de dynamique et ne dépend pas non plus de la quantité d'information. A un niveau plus local, on peut constater que le contenu *Opéra* a été qualifié par les niveaux les plus faibles à la fois de compréhension, de dynamique et de quantité perçue d'information. Ces résultats vont dans le sens de ceux obtenus à partir des séquences évaluées dans le cadre de l'expérimentation B1. Un autre effet notable concerne le contenu *Théâtre* pour lequel le niveau moyen de compréhension était significativement plus faible que pour les autres contenus du corpus ($p < 0{,}001$) hormis *Opéra*. Cet effet n'a pas été reflété par les évaluations réalisées lors de B1 présentant des séquences 2D sans dégradations.

8.9.2.5. CATEGORIE TECHNIQUE

La Figure 8.11 ci-dessous présente le niveau moyen du descripteur *Luminosité* obtenu pour chaque contenu. Une ANOVA considérant la variable indépendante « Contenu » et la variable aléatoire « Participant » et la variable dépendante «Luminosité» à trois modalités (faible, modérée, forte) a indiqué une influence significative du contenu sur le niveau de luminosité évalué par les participants (F (4, 132) = 16,04, p <0,001). Précisément, les contenus *Opéra* et *Sport* étaient significativement plus lumineux (avec p <0, 001) que les contenus *Danse* (le moins lumineux), *Documentaire* (Doc.) et *Théâtre*. La variable « Participant » n'a pas eu d'influence significative sur l'évaluation de ce critère.

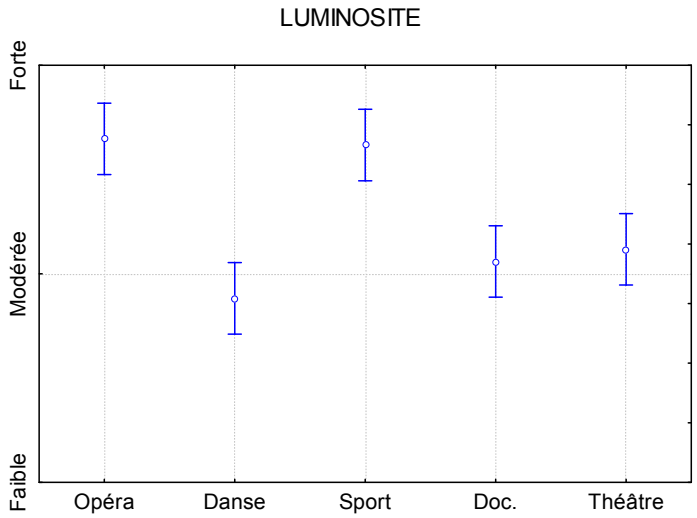


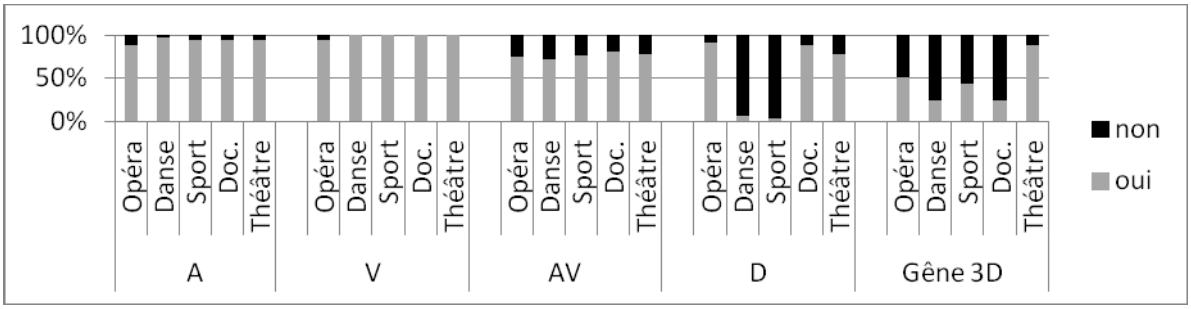
Fig. 8.11. Influence du contenu sur le niveau moyen obtenu pour le descripteur *Luminosité*.

8.9.2.6. CATEGORIE PERCEPTION

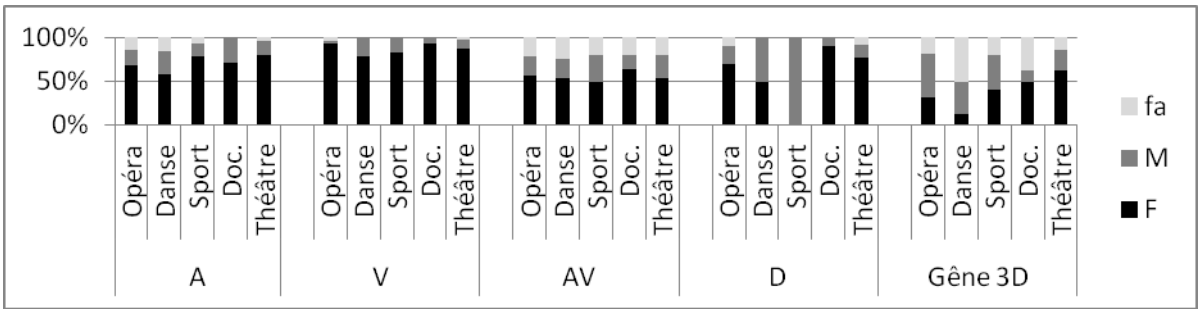
« DETECTION » ET « CERTITUDE »

La Figure 8.12 ci-dessous présente l'influence du contenu sur la détection (répartition par effectifs) de chaque type de dégradations : Audio (A : A-PP), Vidéo (V : V-DEB), AudioVidéo combinées (AV : V-DEB* A-PP), désynchronisation (D : 1500 ms avance du son) et la présence d'une gêne éventuelle liée au format 3D (gêne 3D) ainsi que les niveaux de certitudes associés.

8.12a. Taux de détection (oui/non) pour chaque contenu et chaque type de dégradation.



8.12b. Taux de certitude lors d'une détection (oui).



8.12c. Taux de certitude en absence de détection (non).

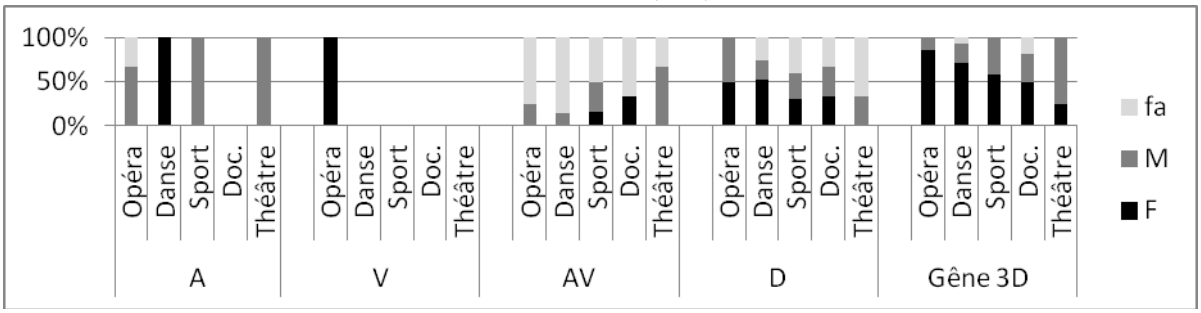


Fig. 8.12. Répartition par pourcentage d'effectifs, pour chaque contenu, du taux de détection (oui/non) des dégradations Audio (A), Vidéo (V), AudioVidéo combinées (AV), Désynchronisation (D) et Gêne 3D et du niveau de certitude associé (faible : fa, Modéré : M ou Fort : F).

Les figures ci-dessus indiquent que les dégradations A, V et AV ont quasiment toujours été détectées avec un niveau élevé de certitude et ce, indépendamment du type de contenu. En revanche, la détection de la désynchronisation et de gêne 3D est dépendante du contenu visualisé. En effet, la désynchronisation n'a clairement pas été détectée pour les contenus *Danse* et *Sport* (respectivement annotés avec les modalités AV et V). Par ailleurs, les participants ont majoritairement rapporté une gêne 3D pour le contenu *Théâtre* avec 87,9% des participants ayant déclaré avoir ressenti de la gêne 3D dont 62,1% avec une certitude forte (pour plus de détail voir l'annexe 8-H qui présente le pourcentage de détection de chaque dégradation pour chaque contenu ainsi que celui du niveau de certitude associé).

Les Figures 8.12b et 8.12c indiquent que globalement la détection d'une dégradation était associée à un niveau de certitude élevé à l'exception de la gêne 3D ayant été à l'origine d'une répartition plus hétérogène des effectifs selon les différents niveaux de certitude. En revanche,

lorsque les dégradations n'étaient pas détectées, les participants étaient moins certains de leurs réponses.

« COMPREHENSION » ET « EMOTIONS NEGATIVES »

La Figure 8.13 ci-dessous présentent les niveaux moyens obtenus pour les descripteurs *Compréhension* (plus exactement dans quelle mesure la dégradation considérée a influencé la compréhension du contenu) et *Emotion négative* pour les dégradations A et AV. En effet, une série d'ANOVAs a été conduite à partir de la variable indépendante « Contenu » et la variable aléatoire « Participant » sur les variables dépendantes « Compréhension» et « Emotions négatives » à cinq modalités (« Pas du tout » à « Extrêmement ») évaluées si et seulement si la dégradation A, V, AV, D ou Gêne 3D étaient détectées. L'ensemble des résultats (effets principaux) obtenus est présenté dans l'annexe 8-I. Ces derniers ont indiqué une influence significative de la variable « Participant » sur la majorité des variables dépendantes. Ils ont également révélé un effet de la dégradation A sur les variables « Compréhension » (F (4, 116) = 17,60, p <0,001) et « Emotions négatives » (F (4, 116) = 7,62, p <0,001) et de la dégradation AV sur la variable « Compréhension » (F (4, 86) = 3,90, p <0,01) comme illustrée dans la figure 8.13.

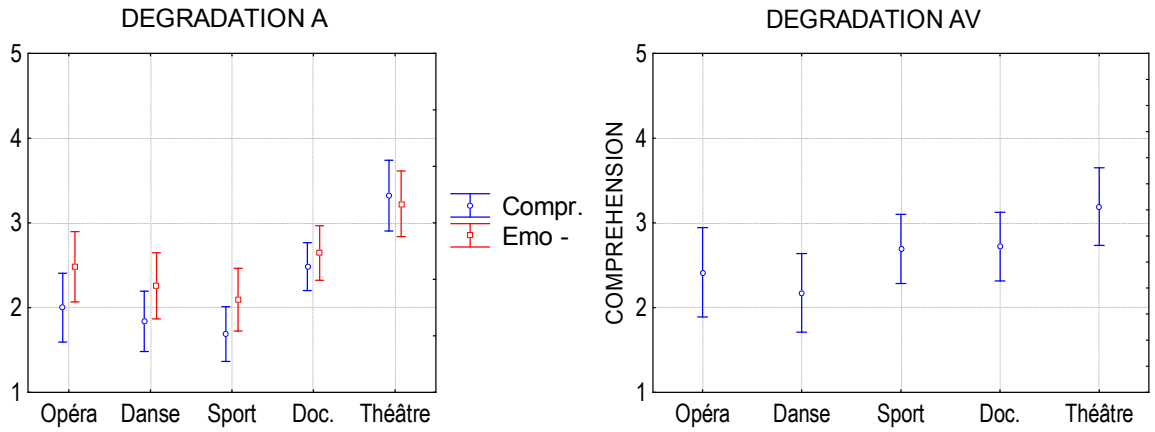


Fig. 8.13. Effet des dégradations audio (A) et AudioVidéo combinée (AV) sur les niveaux de compréhension et d'émotions négatives (Emo -) de 1 (Pas du tout) à 5 (Extrêmement).

Comme cela peut être observé, les participants ont rapporté que la dégradation A a impacté leur niveau de compréhension lorsque celle-ci était appliquée au contenu *Théâtre* (impact de la dégradation sur la compréhension de ce contenu significativement plus fort que pour l'ensemble des autres contenus avec p <0,001 entre Théâtre et Opéra, Danse, Sport et p <0,01 entre Théâtre et Documentaire) et dans une moindre mesure, au contenu *Documentaire* (impact de la dégradation sur la compréhension de ce contenu significativement plus fort que pour les contenus Danse avec p <0,05 et Sport avec p <0,01).

La dégradation AV a, quant à elle, impacté plus fortement la compréhension du contenu *Théâtre* par rapport aux contenus *Opéra* (p < 0,05) et *Danse* (p < 0,01).

« QUALITE PERÇUE »

La Figure 8.14 ci-dessous présente les notes MOSAV, MOSV et MOSA obtenues pour chaque contenu.

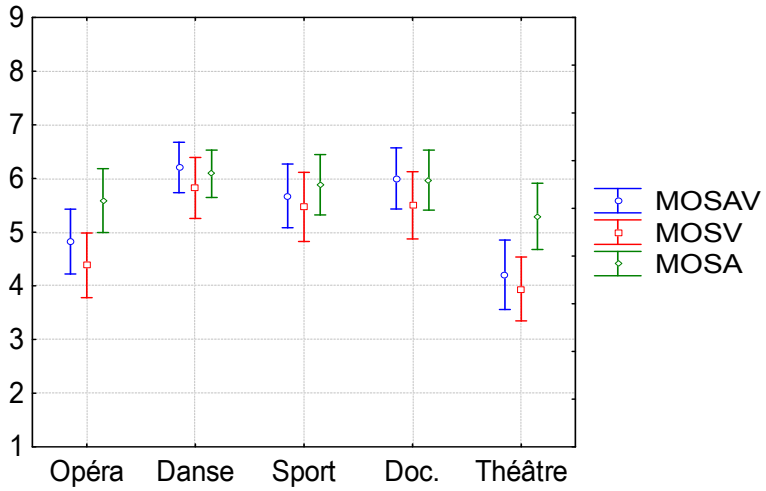


Fig. 8.14. Effet du contenu sur MOSAV, MOSV et MOSA.

L'observation de la figure indique que la qualité restituée est globalement satisfaisante. Néanmoins, les dégradations apparaissant sur les contenus *Opéra* et *Théâtre* ont plus fortement altéré la perception des qualités vidéo et audiovisuelle. Ce constat est confirmé par une ANOVA considérant la variable indépendante « Contenu » et la variable aléatoire « Participant » sur les variables dépendantes « QAV », « QV » et « QA ». Les résultats ont en effet révélé une influence significative du contenu sur QAV (F (4, 132) = 15,18, p <0,001) et QV (F (4, 132) = 11,51, p <0,001). En revanche, l'évaluation de QA n'a pas été influencée par le contenu (F (4, 132) = 2,33, p =0,06). Un effet de la variable « Participant » a également été constaté sur QAV (F (32, 132) = 4,67, p <0,001), QV (F (32, 132) = 3,74, p <0,001) et QA (F (32, 132) = 4,47, p <0,001)

Plus précisément, une diminution des scores MOSAV et MOSV pour les contenus *Opéra* et *Théâtre* a été révélée : MOSAV est significativement plus basse pour *Théâtre* que pour l'ensemble des autres contenus, à l'exception d'Opéra, (avec p <0,001 entre Théâtre et Danse, Documentaire et p <0,01 entre Théâtre et Sport) et pour *Opéra* comparativement aux contenus *Danse* (avec p <0,01) et *Documentaire* (avec p <0,05). MOSV est aussi significativement plus faible pour les contenus *Opéra* et *Théâtre* par rapport aux autres contenus du corpus (avec p < 0,001 entre Opéra et Danse, p <0,01 entre Opéra et Documentaire, p <0,05 entre Opéra et Sport et avec p <0,001 entre Théâtre et Danse, Sport, Documentaire). La figure permet aussi de constater que QV tend à attirer QAV comme l'illustre les notes attribuées aux contenus *Opéra* et *Théâtre*.

8.9.3. CONCLUSIONS MESURES SUBJECTIVES

L'analyse des données subjectives a tout d'abord révélé un **impact de la passation sur le niveau de fatigue**, plus important après le test. La visualisation de contenus présentés en format 3D et comportant des dégradations audio et/ou vidéo ainsi que l'activité de complétion de questionnaires ont donc influencé, comme cela pouvait être attendu, le niveau de fatigue subjectivement évalué. Cette information peut être pertinente dans le cadre de l'interprétation des mesures psychophysiologiques.

Le classement des contenus par ordre de préférence avant le test a permis de constater que globalement les participants n'avaient pas d'*a priori* sur les contenus à visualiser, c'est-à-dire une absence de rejet ou au contraire d'intérêt fort pour un ou plusieurs contenus qui auraient pu biaiser les résultats. En revanche, après visualisation, le classement a dévoilé un rejet massif du contenu *Opéra* tandis qu'à l'inverse, *Danse* a été le contenu préféré par les participants.

Globalement, la caractérisation, à partir des descripteurs évalués après la visualisation des contenus entiers, a permis de retrouver celle réalisée à partir des courtes séquences non dégradées (moins vrai pour les descripteurs Modalité et Dynamique). Ainsi, **la procédure de caractérisation à partir d'extraits de quelques secondes semble être représentative des contenus entiers** principalement concernant les descripteurs de la catégorie Hédonique. Le rejet du contenu *Opéra* a donc été observé indépendamment de la présence de dégradations ou du type de format de présentation (2D ou 3D). Cette information devra être prise en compte dans la suite des analyses. Il en sera de même pour le contenu *Danse* ayant suscité une forte adhésion (pattern inverse à Opéra).

L'évaluation de la qualité audio, vidéo et audiovisuelle a révélé que la qualité restituée était globalement satisfaisante. Néanmoins, les contenus ***Opéra* et *Théâtre* ont reçu des notes de qualité vidéo et audiovisuelle plus faibles que pour les autres contenus**. Dans l'expérimentation A, *Opéra* avait obtenu, à l'inverse, des notes de qualité vidéo et audiovisuelle plus élevées que la majorité des autres contenus. Plusieurs explications peuvent être apportées pour éclaircir ce point. Tout d'abord, ce résultat pourrait traduire un effet positif du débit adaptatif (ainsi, l'effet observé lors de l'étude A s'expliquerait plus par un effet inégal d'un débit fixe que par un effet de la modalité dominante). Par ailleurs, le contenu *Opéra* s'est révélé être un contenu très peu apprécié des participants tant en matière d'intérêt, de plaisir que d'arousal, il a également été annoté par les niveaux les plus faibles de compréhension, de dynamique et de quantité d'information perçue. L'évaluation de ces critères (jugés après la qualité mais présentés avant la passation de test) aurait pu conduire les participants à homogénéiser leurs réponses sur l'ensemble du questionnaire. Ainsi, les notes de qualité obtenues pourraient avoir été « contaminées » notamment par l'évaluation des niveaux d'intérêt, de plaisir, *etc.* très faibles pour ce contenu en particulier. Cet effet contaminant (maintien d'une cohérence inter-critères) pourrait être rapproché de l'influence positive observée par Palhais *et al.* (2012) du niveau d'intérêt sur l'évaluation subjective de la

qualité vidéo. Les résultats de l'expérimentation C indiquent que l'intérêt (lorsque celui-ci est subjectivement évalué), et plus largement la qualité hédonique d'un contenu, a aussi une influence négative sur l'évaluation subjective de la qualité vidéo et audiovisuelle (la qualité audio n'a pas été influencée par le type de contenu).

La présentation 3D pourrait également expliquée la diminution des notes de qualité vidéo et audiovisuelle constatée pour le contenu *Théâtre*. En effet, les participants ont rapporté avoir été gênés par le format 3D du contenu *Théâtre*, cet effet pouvant être expliqué par la qualité native 3D de ce contenu probablement moins bonne que celle des autres contenus (présence de *ghosting*, défauts de luminance).

Par ailleurs, les notes de qualité audiovisuelle ont plus largement reflété les notes de qualité vidéo qu'audio (fig. 8.14) indépendamment de la nature de la modalité dominante. Ce constat permet de croire que, dans le cadre de cette expérimentation, la qualité vidéo était prédominante lors de l'élaboration de la note de qualité audiovisuelle. Dans l'expérimentation A, les qualités audio et vidéo semblaient contribuer de manière plus ou moins équivalente à la qualité perçue audiovisuelle, cette dernière correspondant globalement à une moyenne des notes de qualité audio et vidéo. Ainsi, **la prédominance de la qualité d'une modalité sur la qualité audiovisuelle perçue semble dépendre d'un certain nombre de facteurs qui ne seraient pas seulement propres au contenu mais aussi au type de critères évalués** (intérêt, plaisir, *etc.*) **ou plus largement au protocole appliqué** (format de présentation : 2D/3D, durée des séquences, durée et seuil des dégradations, nombre de présentation des séquences, *etc.*). Ce constat souligne l'importance de demander systématiquement aux participants d'évaluer les qualités audio et vidéo séparément au lieu de la seule qualité audiovisuelle comme cela est actuellement recommandé par la norme UIT-T P.911 (UIT-T, 1998).

Sur le plan de la perception des dégradations de qualité, **les dégradations audio, vidéo et audio-vidéo** (combinaison des dégradations audio et vidéo) **ont été détectées avec une certitude forte** par la majorité des participants et ce, indépendamment du type de contenu. En revanche, la **désynchronisation n'était pas ou peu perçue pour les contenus *Danse* et *Sport***. Cela peut être expliqué par le fait que le contenu *Sport* comportait de nombreux commentaires extra-diégétiques et que le contenu *Danse* présentait très peu de scènes verbales. A l'inverse, les contenus *Opéra*, *Documentaire* et *Théâtre* présentaient principalement des sons de parole diégétiques. L'expérience B2 a montré que pour être perceptible la désynchronisation doit survenir sur des scènes auditives à la fois verbales et diégétiques. Ainsi, l'absence de détection de la désynchronisation pour les contenus *Danse* et *Sport* peut s'expliquer par le fait que ces contenus ne remplissaient pas les conditions nécessaires à la perception de cette dégradation. Ce constat confirme une nouvelle fois l'importance de la caractérisation préalable des séquences de test. Il souligne également la **pertinence de l'ajout d'une question spécifique à la détection des dégradations** aux questionnaires d'évaluation de la qualité audiovisuelle, notamment lorsque la dégradation par désynchronisation image/son souhaite être étudiée.

Par ailleurs, l'évaluation de la présence d'une gêne liée au format 3D n'était pas associée à une certitude forte. Il semble que l'évaluation de ce critère soit plus amplement soumise à la

sensibilité individuelle des participants. De manière générale, lorsqu'une dégradation n'était pas détectée, les participants n'étaient pas certains de leurs réponses. Ce constat pourrait être expliqué par la présence même d'une question sur la détection d'une dégradation donnée sous-tendant la présence possible de ce type de dégradation (biais lié au questionnaire).

Dans cette étude, l'impact des dégradations sur la compréhension des participants et sur le sentiment d'émotions négatives (frustration, agacement, *etc.*) a également été évalué. Les résultats portant sur ces critères ont montré plusieurs effets intéressants. Tout d'abord, **la dégradation audio a eu l'impact le plus important sur la compréhension et le sentiment d'émotions négatives** lors de son application aux contenus *Théâtre* et *Documentaire*. Ces deux contenus ont en commun la présence de nombreuses scènes verbales et diégétiques. La dégradation audio survenant lors de la visualisation de ces contenus entraînerait une perte d'intelligibilité gênante pour la compréhension du contenu. Pour autant, cela n'est pas observé pour le contenu *Opéra* également caractérisé par un contenu verbal et diégétique et ayant été annoté avec un niveau faible de compréhension générale. Cependant, il s'agissait exclusivement de paroles chantées en langue étrangère. Par conséquent, l'accès au sens (déterminé par les sons de parole pour Théâtre et Documentaire) serait moins déterminé pour *Opéra* par la compréhension des sons de parole que par les informations musicales et visuelles interprétées dans leur ensemble. Par ailleurs, l'effet de la dégradation audio sur la compréhension de *Documentaire* n'a pas été reflété par l'évaluation du niveau de compréhension globale du contenu, contrairement à *Théâtre*. Cela peut être expliqué par le fait que ce contenu présentait un certain nombre de séquences à dominance vidéo (c.-à-d. pour lesquelles le contenu audio n'était pas verbal) contrairement à *Théâtre* pour lequel quasiment toutes les séquences étaient verbales. *Documentaire* a d'ailleurs été qualifié par une modalité AV mais avec une tendance vidéo par les participants et par une modalité dominante vidéo selon la caractérisation experte.

L'évaluation du sentiment d'émotions négatives consécutives à la dégradation audio semble refléter les notes attribuées à la compréhension, toujours évaluée avant. Soit la perte de compréhension est à l'origine d'émotions négatives, soit l'ordre de présentation des questions influence les réponses des participants qui essaient de maintenir une certaine cohérence entre leurs évaluations aux différents critères. Toutefois, il est possible que ces deux suppositions ne s'excluent pas l'une l'autre.

La dégradation audio-vidéo a également impacté le niveau de compréhension du contenu *Théâtre* (par rapport au contenu Danse et Opéra). Cet effet peut être expliqué par l'influence de la dégradation audio comme discuté ci-avant (perte d'intelligibilité).

La dégradation audio a donc diminué la *qualité d'expérience* du spectateur, pour les contenus *Théâtre* et *Documentaire* en particulier, **sans pour autant avoir été reflétée par l'évaluation de la qualité audio**. Ce constat attire l'attention sur la limite de l'évaluation seule des niveaux de qualité perçue qui ne permet donc pas de rendre compte de l'ensemble des influences de la qualité sur la *qualité d'expérience* du spectateur. Ce résultat renforce la nécessité de proposer un questionnaire enrichi, et plus généralement une méthode

d'évaluation holistique, pour étendre l'évaluation de la qualité du signal restitué à l'évaluation de la *qualité d'expérience*.

Le rejet du contenu *Opéra*, l'adhésion pour le contenu *Danse* ou l'effet de la présentation 3D et de la dégradation audio sur le contenu *Théâtre* sont autant d'informations qui pourraient permettre une meilleure interprétation des mesures psychophysiologiques.

8.9.4. MESURES OCULAIRES

8.9.4.1. REDUCTION DES DONNEES

Pour chacun des dix-neuf participants sélectionnés (sect. 8.9.1 ci-dessus) environ une heure trente de données ont été recueillies et ce, pour chaque indice oculaire. Sur la base des informations de synchronisation obtenue *via* le module *Télécommande*, une première étape a consisté à faire correspondre les mesures enregistrées au déroulé de la passation (baseline, amorce, contenus) ainsi qu'à celui de chaque contenu avec ses périodes dégradées et non dégradées.

Chaque contenu était divisé en huit (Danse, Théâtre, Sport) ou neuf périodes (Documentaire, Opéra) dont quatre, d'une durée d'une minute, qui présentaient une dégradation Audio (A), Vidéo (V) AudioVidéo combinée (AV) ou Désynchronisation (D). Les périodes sans dégradations étaient notées P1, P2, P3, P4 et éventuellement P5 selon leur ordre chronologique d'apparition. Deux types d'analyses ont été effectués :

- le premier considère, pour chaque participant, la **moyenne temporelle des signaux calculée pour l'amorce et chacun des cinq contenus de test** (soit 6 scalaires par participant). Pour des raisons identiques à celles de l'expérimentation A (difficulté de contraindre le participant à fixer un écran noir durant plusieurs minutes), les données oculaires enregistrées durant la baseline n'ont pas été prises en compte. Les données oculaires n'étaient donc pas normalisées,
- le deuxième considère, pour chaque participant et pour chaque contenu, **la moyenne temporelle calculée pour chacune des périodes dégradées et non dégradées** d'un contenu donné (soit 42 scalaires par participant).

Ces deux niveaux d'analyses devaient permettre l'étude des influences du type d'*activité* (amorce *vs.* premier contenu visualisé), du *contenu* (Danse, Documentaire, Opéra, Sport, Théâtre) et de la *période* (dégradée et non dégradée).

8.9.4.2. EFFET DU TYPE D'ACTIVITE

Une série de tests de *Student* réalisée pour chaque indicateur entre les moyennes obtenues pour l'amorce et celle du premier contenu visualisé n'a pas révélé de différences significatives : DP ($t(18) = 1,45$, $p = 0,17$), EBdur ($t(18) = 0,66$, $p = 0,52$), EBfreq ($t(18) = 0,17$, $p = 0,87$), PERCLOS ($t(18) = -0,50$, $p = 0,62$) et SAC ($t(18) = -0,93$, $p = 0,37$).

8.9.4.3. EFFET DU CONTENU

La Figure 8.15 ci-dessous présente les moyennes obtenues pour les indices DP et SAC pour chacun des contenus visualisés. Elle illustre l'effet significatif du contenu sur ces deux indices oculaires seulement, conformément aux ANOVAs réalisées à partir de la variable indépendante « Contenu » et aléatoire « Participant » et des variables dépendantes « DP », « EBdur », « EBfreq », « PERCLOS », « SAC ». Les résultats des ANOVAs ont en effet montré une influence significative du contenu uniquement pour les indices SAC : $F(4, 72) = 5{,}56$, $p < 0{,}001$ et DP : $F(4, 72) = 8{,}13$, $p < 0{,}001$. Un effet systématique de la variable « Participant » a été observé avec $p < 0{,}05$ pour SAC et $p < 0{,}001$ pour tous les autres indicateurs. L'ensemble des résultats (effets principaux) est présenté dans l'annexe 8-J_A.

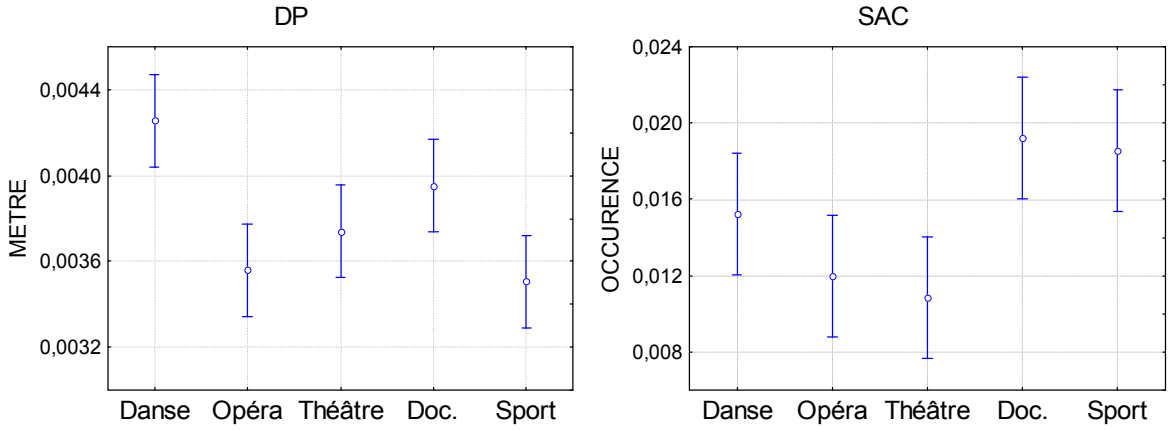


Fig. 8.15. Moyennes obtenues pour chaque contenu pour les indices DP et SAC (0 = absence de saccades et 1= présence de saccades).

Plus précisément, comme l'indique la figure, une augmentation significative est observée pour DP lors de la visualisation du contenu *Danse* par rapport aux contenus *Opéra*, *Sport* (avec $p < 0{,}001$) et *Théâtre* (avec $p < 0{,}05$). Une augmentation de l'indice SAC a également été constatée lors de la visualisation des contenus *Documentaire* et *Sport* par rapport aux contenus *Opéra* et *Théâtre* (avec $p < 0{,}05$).

8.9.4.4. EFFET DES DEGRADATIONS

L'hypothèse principale (H0p) supposait que la présence de dégradations audio et/ou vidéo (désynchronisation, réduction du débit) pourrait être à l'origine d'un effort mental (H1p) supplémentaire (pour décoder et interpréter le message dégradé) puis d'un état de fatigue (H2p), visibles à travers des modifications des patterns physiologiques et/ou oculaires. Pour étudier cela, une ANOVA a été réalisée pour chaque contenu, en considérant la variable indépendante « Période » (P1, P2, P3, P4, éventuellement P5 et A, V, AV, D) et la variable aléatoire « Participant » et les variables dépendantes « DP », « EBdur », « EBfreq », « PERCLOS », « SAC ». Un effet de la variable « Participant » a été trouvé pour tous les indicateurs étudiés et ce pour chaque contenu (excepté l'indicateur SAC pour le contenu

Opéra). L'ensemble des résultats (effets principaux) est présenté dans l'annexe 8-J_B. Les effets significatifs de la période sont présentés dans le Tableau 8.5 ci-après.

Tableau 8.5. Effets significatifs de la variable indépendante (VI) « Période » en considérant la variable aléatoire « Participant » sur les variables dépendantes (VD) « EBdur », « SAC » et « DP » étudiées pour chaque contenu Danse, Documentaire, Opéra, Sport et Théâtre.

VI	Contenus	VD	Somme des carrés	ddl effet	ddl erreur	Moyenne des carrés	F	*p*
Périodes	**Danse**	**EBdur**	0,09	7	126	0,01	10,63	<0,001
		DP	0,00	7	126	0,00	9,78	<0,001
	Documentaire	**EBdur**	0,05	8	144	0,00	3,59	<0,001
		DP	0,00	8	144	0,00	10,24	<0,001
	Opéra	**EBdur**	0,09	8	144	0,02	5,55	<0,001
		DP	0,00	8	144	0,00	10,58	<0,001
	Sport	**EBdur**	0,05	7	126	0,00	6,53	<0,001
		DP	0,00	7	126	0,00	16,91	<0,001
	Théâtre	**EBdur**	0,03	7	126	0,00	2,98	<0,01
		SAC	0,00	7	126	0,00	3,03	<0,01
		DP	0,00	7	126	0,00	15,60	<0,001

Globalement, les résultats ont indiqué un effet de la période sur EBdur et DP pour l'ensemble des contenus. Comme illustré dans Figure 8.16 pour le contenu *Sport*, l'effet constaté sur ces deux indicateurs peut être expliqué par un effet du début du contenu circonscrit à la première période du contenu (non dégradée -P1- ou dégradée -A- pour Danse) significativement inférieure à la majorité des autres périodes constituant un contenu donné. Ainsi, le début du contenu semble avoir influencé ces deux indicateurs indépendamment de la présence de dégradations.

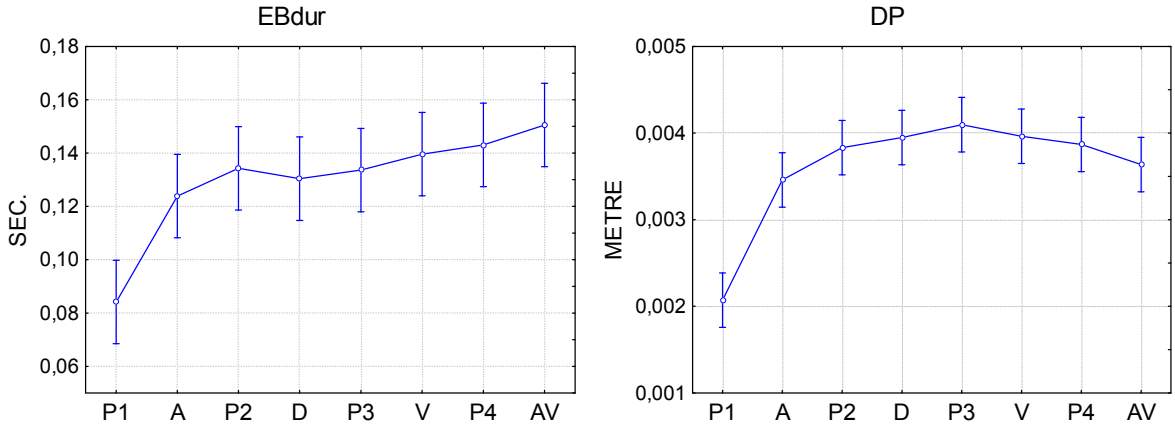


Fig. 8.16. Moyennes obtenues pour chaque période du contenu Sport pour les indices EBdur et DP. P1, P2, P3 et P4 correspondent aux périodes non dégradées (selon chronologie du contenu). A correspond à la période présentant la dégradation audio, V à la dégradation vidéo, D à la dégradation désynchronisation et AV à la combinaison des dégradations A et V, selon leur ordre d'apparition.

Le Tableau 8.5 ci-dessus indique également une influence de la période sur l'indicateur SAC lors de la visualisation du contenu *Théâtre* (voir fig. 8.17 ci-dessous). Cet effet s'explique par une différence significative entre la période P4 et les périodes P1 (avec p <0,01), P2 et AV (avec p <0,05).

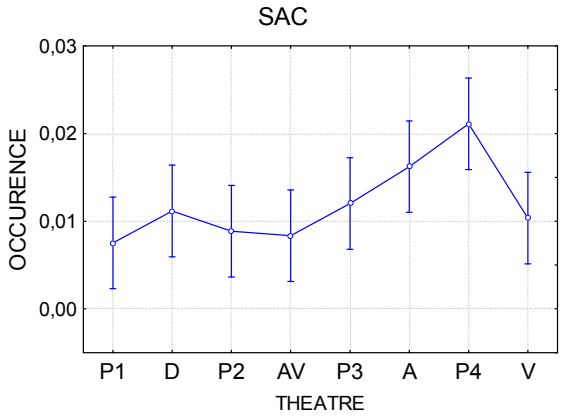


Fig. 8.17. Moyennes obtenues pour chaque période du contenu Théâtre pour l'indice SAC. P1, P2, P3 et P4 correspondent aux périodes non dégradées (selon chronologie du contenu). A correspond à la période présentant la dégradation audio, V à la dégradation vidéo, D à la dégradation désynchronisation et AV à la combinaison des dégradations A et V, selon leur ordre d'apparition.

8.9.5. CONCLUSIONS MESURES OCULAIRES

Aucune différence significative entre le contenu *amorce* et la visualisation du premier contenu n'a été observée. Ainsi, si le début de l'activité de visualisation génère une variation des indicateurs oculaires (ce qui ne peut être vérifié car aucune baseline ne peut être mesurée pour les indicateurs oculaires), le contenu *amorce* a bien rempli le rôle d' « absorber » l'effet lié au changement d'activité.

L'étude de l'activité oculaire a révélé un effet marqué du contenu sur les variations observées. Premièrement, le contenu a influencé le diamètre pupillaire pour lequel une augmentation a été constatée lors de la visualisation du contenu *Danse*. Cet effet peut en partie être expliqué par le niveau de luminosité. Comme l'indique le Tableau 8.6 ci-après, le contenu *Danse* correspondait à l'un des extraits les moins lumineux du corpus audiovisuel, tel que rapporté par la caractérisation experte. Ainsi, la dilatation pupillaire observée peut être attribuée à un réflexe photo-moteur. A l'inverse le contenu *Sport*, pour lequel la taille du diamètre pupillaire était la plus faible correspondait au contenu le plus lumineux.

Tableau 8.6. Moyennes (calculées à partir de l'ensemble des séquences annotées d'un contenu donné) obtenues pour le descripteur Luminosité (où 1 correspond à un niveau « faible » et 3 à un niveau « fort ») selon la caractérisation experte réalisée pour la totalité des contenus du corpus.

Contenus	Luminosité
Sport	3
Opéra	1,82
Documentaire	1,43
Danse	1,08
Théâtre	1

En revanche, la caractérisation experte indique également que le contenu le moins lumineux est le contenu *Théâtre*. Selon l'explication liée au réflexe pupillaire, *Théâtre* aurait dû être à l'origine d'une dilatation plus importante de la pupille que le contenu *Danse*. Or, la moyenne de diamètre pupillaire obtenue pour *Théâtre* était inférieure à celle obtenue pour le contenu *Danse* de façon significative. Cette observation invite à chercher une explication autre que l'influence seule du niveau de luminosité. L'analyse de l'expérience subjective a révélé que le contenu *Danse* avait été jugé par les participants comme le contenu le plus intéressant, le plus plaisant et suscitant le plus fort arousal. Or, le diamètre pupillaire augmente lors d'une émotion positive et lors d'une augmentation du niveau d'arousal (sect. 4.1, chap. IV). Ainsi, **l'augmentation constatée pour le contenu *Danse* pourrait traduire une activation du système nerveux sympathique (SNS) qui ne serait pas uniquement expliquée par un réflexe photo-moteur mais aussi par la valence positive et le niveau élevé d'arousal caractérisant ce contenu.** A titre indicatif, l'effet du facteur « Sexe » a été vérifié et aucune différence Homme/Femme n'a été révélée par cette analyse[25].

Le nombre de saccades a également varié en fonction du type de contenu, en effet, un plus grand nombre de mouvements oculaires a été observé lors de la visualisation des contenus *Documentaire* et *Sport*. Le niveau élevé de *dynamique caméra* tel que caractérisé par l'expert (voir tabl. 8.7 ci-dessous) pourrait être une explication valable. En effet, *Documentaire* et *Sport* ont été qualifiés par les niveaux les plus forts de *dynamique caméra* (c.-à-d. nombreux changements de plans/scènes, de décors, de recadrages, travellings, *etc.*). A chaque mouvement de caméra, la zone visuelle d'intérêt pourrait être déplacée obligeant le participant à de nombreux réajustements oculaires et à une poursuite visuelle plus intense (suivi des personnages, des scores affichés, *etc.*) pour maintenir son attention sur les zones pertinentes de l'image. A l'inverse, le nombre de saccades était significativement moins important pour les contenus *Opéra* et *Théâtre* caractérisés par les niveaux les plus faibles de *dynamique caméra* (plan fixe, très peu de changements de plans, de décors, de recadrages, *etc.*). Ces deux derniers contenus présentaient donc un contexte fortement statique ne nécessitant pas de nombreux déplacements oculaires (action essentiellement localisée au centre de l'écran).

[25] ANOVA considérant les variables indépendantes « Sexe » et « Contenu » et la variable dépendante « DP ».

Ainsi, l'effet observé sur l'indicateur de saccades peut raisonnablement être expliqué par un effet du niveau de dynamique de la caméra.

Tableau 8.7. Moyennes (calculées à partir de l'ensemble des séquences annotées d'un contenu donné) obtenues pour les descripteurs Dynamique de contenu et Dynamique caméra (notés de 1 : « faible » à 3 : « forte») selon la caractérisation experte réalisée pour la totalité des contenus du corpus.

Contenus	Dynamique Contenu	Dynamique caméra
Sport	1,95	1,73
Opéra	1,76	1
Doc.	1,59	1,37
Danse	2,04	1,16
Théâtre	1,63	1,04

L'étude de la période (dégradée et non dégradée) devait permettre de faire émerger un éventuel effet de la qualité sur les mesures oculaires. Cependant, l'analyse a seulement mis en avant un effet du début du contenu visualisé. En effet, la taille du diamètre pupillaire et la durée de clignement de l'œil augmentaient significativement après la toute première période (dégradée ou non) des contenus. L'effet sur le diamètre pupillaire pourrait exprimer une activation du SNS après le début du contenu (la première période était d'une durée variable de 55 s et 4 min 20, selon le contenu) liée à un phénomène attentionnel (attention accrue, processus de traitement de l'information) résultant de l'engagement du participant dans l'activité de visualisation. L'effet observé pourrait aussi refléter le changement de luminosité entre l'écran noir (durant la complétion ainsi que durant le décompte) et l'affichage des premières scènes d'un contenu donné. Ainsi, la diminution du DP observé au début du contenu correspondrait à un réflexe photo-moteur suivi par une adaptation de la pupille au niveau de luminosité du contenu. La durée de clignement, plus faible en début de contenu, pourrait être interprétée comme un phénomène attentionnel reflétant l'engagement du participant dans l'activité de visualisation.

Un effet de la période sur l'indicateur de saccades a également été mis en évidence pour le contenu *Théâtre*. L'effet significatif s'explique par une augmentation du nombre de saccades pour une des périodes, non dégradée, du contenu (P4, voir fig. 8.17). Toutefois, il semblerait que l'effet intéressant pour ce contenu vienne de la rupture de la courbe observée lors de l'introduction de la dégradation vidéo : le nombre de saccades chutait alors, tandis qu'une évolution constante peut être observée pour le reste du contenu. Un effet de la dégradation vidéo peut être supposé (la dynamique n'augmentant pas pour cette période). Celle-ci était introduite lors de la dernière période du contenu. La dégradation vidéo, présentée à la fin du contenu, cumulée à un effet néfaste du format 3D (rappelons que ce contenu a été noté avec les niveaux de qualité vidéo et audiovisuelle les plus faibles), aurait pu conduire à un effort mental suffisant pour générer un état de fatigue exprimé par l'indicateur de saccades. Cependant, cet effet potentiel n'est pas appuyé par d'autres indicateurs. Il est aussi possible que l'effet observé pour *Théâtre* relève d'une influence propre au contenu mais non précisée par les descripteurs étudiés (dynamique, nombre de personnages, *etc.*).

Les résultats obtenus n'ont donc pas permis d'observer un impact des fluctuations de qualité sur les indicateurs oculaires étudiés. Ce constat infirme les hypothèses, principales et appliquées, soutenant une modification de l'activité oculaire du spectateur en réaction à la présence de dégradations audio et/ou vidéo.

Notons que les analyses réalisées n'ont pas exprimé la gêne éprouvée vis-à-vis de la présentation 3D du contenu *Théâtre*. Une réponse des indicateurs de fatigue (↑PERCLOS, EBdur, EBfreq, ↓ SAC) pour ce contenu pouvait être attendue. De manière générale, la fatigue consciemment ressentie après le test n'a pas été exprimée du point de vue du comportement oculaire. Trois explications sont possibles :

- **un effet biaisant de l'évaluation subjective** de la fatigue sur la réponse des participants : l'évaluation du niveau de fatigue après le test conduirait le participant à supposer que son niveau de fatigue a évolué (traduction d'un biais de conformité aux attentes de l'expérimentateur, sect. 1.6.1, chap. I),
- **un effet « possible » :** un état de fatigue consciemment ressenti pourrait ne pas être exprimé physiologiquement,
- **un effet de l'analyse** : un état de fatigue a consciemment été ressenti et exprimé physiologiquement mais non observé à partir du protocole d'analyse réalisé.

8.9.6. MESURES PHYSIOLOGIQUES

La température de la salle de test a été mesurée tout au long des passations. Les mesures recueillies ont indiqué que la température de la salle de test était comprise entre 23 et 25,9 C°, pour une moyenne de 24,34 C°. Par ailleurs, les mesures d'AED étant particulièrement sensibles aux variations de température, une analyse bivariée par étude des corrélations de Bravais-Pearson a permis de s'assurer que l'AED n'a pas été influencée par la température de la salle de test ($R=0,13$, $R^2=0,018$). Cette analyse a été conduite à partir des moyennes de température de la salle de test et de la conductance cutanée, obtenues pour chacune des périodes d'un contenu donné.

8.9.6.1. REDUCTION DES DONNEES

Pour chacun des vingt-six participants sélectionnés environ une heure trente de données ont été recueillies et ce, pour chaque indice physiologique. Comme pour les mesures oculaires, une première étape a consisté à faire correspondre les mesures enregistrées au déroulé de la passation (baseline, amorce, contenus) ainsi qu'à celui de chaque contenu (périodes dégradées et non dégradées). Deux types d'analyses ont été effectués :

- le premier considère, pour chaque participant et pour chaque contenu, **la moyenne temporelle des signaux calculée pour la baseline, l'amorce et chacun des cinq contenus de test** (soit 7 scalaires par participant). Les données physiologiques étaient normalisées,

- le deuxième considère, pour chaque participant et pour chaque contenu, **la moyenne temporelle calculée pour chacune des périodes dégradées et non dégradées** d'un contenu donné (soit 42 scalaires par participant).

Ces deux niveaux d'analyses devaient permettre l'étude des influences du type d'*activité* (baseline, amorce, premier contenu visualisé), du *contenu* (Danse, Documentaire, Opéra, Sport, Théâtre) et de la *période* (dégradée et non dégradée).

Dans l'intention de minimiser la forte variabilité interindividuelle, les données physiologiques ont été normalisées selon deux manières différentes :

(a)

$$\text{signal}n = \frac{\text{Moyenne temporelle } \textit{signal} \text{ (contenu ou période)} - \text{Moyenne temporelle } \textit{bsl}}{\text{Moyenne temporelle } \textit{bsl}}$$

(b)

$$\text{signal}n = \frac{\text{Moyenne temporelle } \textit{signal} \text{ (contenu ou période)} - \text{Moyenne temporelle } \textit{amorce}}{\text{Moyenne temporelle } \textit{amorce}}$$

La première formule est identique à celle réalisée lors de l'expérience A et correspond à une normalisation par la baseline (mesures au repos) et la seconde repose sur une normalisation par l'amorce (vanilla baseline). Selon Jennings *et al.* (1992), cette dernière méthode de normalisation devrait permettre de pallier les problèmes liés à une baseline mesurée au repos, comme l'expression du changement de l'état inactif à la phase d'activité, mais aussi les défauts de vigilance du participant (somnolence, anxiété, ennui, pensées plus ou moins positives) et enfin, d'activer les processus qui seront engagés dans la tâche principale (visualisation audiovisuelle 3D ici). Ces deux approches devaient permettre d'étudier la méthode de normalisation la plus adéquate. Les données normalisées seront notées de la manière suivante : VSP*n*, FC*n*, AED*n* et TCP*n*.

8.9.6.2. EFFET DU TYPE D'ACTIVITE

L'effet du changement d'activité (baseline, amorce et premier contenu visualisé -C1-) et la pertinence de l'amorce ont tout d'abord été étudiés. Pour cette première étape seulement, les données physiologiques n'étaient pas normalisées afin de pouvoir comparer les moyennes obtenues pour chaque contenu, pour l'amorce et pour la baseline.

La Figure 8.18 ci-après illustre l'influence de chaque type d'activité sur les niveaux moyens d'AED et de FC.

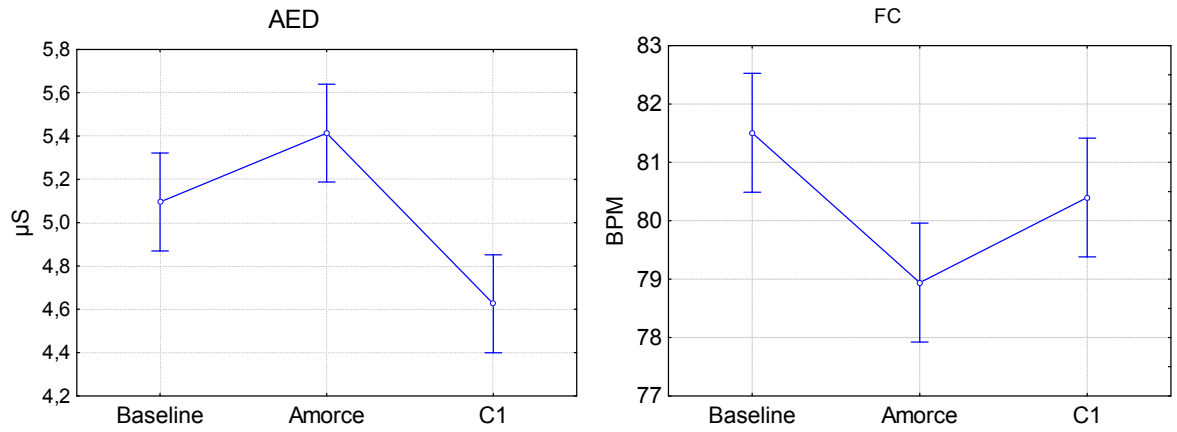


Fig. 8.18. Moyennes d'AED et de FC obtenues pour chaque activité : Baseline, Amorce et C1.

La Figure 8.18 permet d'observer des patterns inverses entre les variations d'AED et de FC. Les résultats d'une ANOVA réalisée pour chacun des quatre indicateurs physiologiques (variables dépendantes : « AED », « FC », « VSP » ou « TCP») avec pour variables indépendantes l'« Activité » (baseline, amorce ou C1) et le « Participant » (aléatoire) ont révélé que la variable « Activité » a influencé l'ensemble des indicateurs : AED (F (2, 50) = 4,09, p <0,001), FC (F (2, 50) = 6,45, p <0,01), TCP (F (2, 50) = 6,74, p <0,01) à l'exception du VSP (F (2, 50) = 1,0, p = 0,57). Un effet de la variable « Participant » a également été trouvé pour l'ensemble des indicateurs (AED : F (25,50) = 108,68, p <0,001), FC (F (25,50) = 71,1, p <0,001), TCP (F (25,50) = 91,43, p <0,001) et VSP (F (25,50) = 2,0, p <0,01).

Plus précisément, les niveaux moyens d'AED étaient significativement plus élevés pour la baseline et l'amorce que lors du premier contenu (p <0,05 entre C1 et Baseline et p < 0,001 entre C1 et Amorce). En revanche, la FC diminue de manière significative lors de la présentation de l'amorce (p <0,001 entre Amorce et Baseline). Enfin, l'effet de la variable « Activité » sur les niveaux moyens de TCP a indiqué une augmentation significative de C1 par rapport à la baseline uniquement (p < 0,01).

8.9.6.3. EFFET DU CONTENU

La Figure 8.19 ci-dessous présente les moyennes de l'AED*n* obtenues pour chacun des contenus visualisés, l'AED*n* étant le seul indice significativement influencé par le contenu (F (4,100) = 3,42, p <0,05) selon une série d'ANOVAs, considérant les variables indépendantes « Contenu » et « Participant » -aléatoire- et les variables dépendantes « AED*n* », « FC*n* », « VSP*n* », « TCP*n* ». Une influence significative de la variable « Participant » a été observée pour les quatre indicateurs étudiés avec p < 0,001. L'ensemble des résultats (effets principaux) est présenté dans l'annexe 8-K_A.

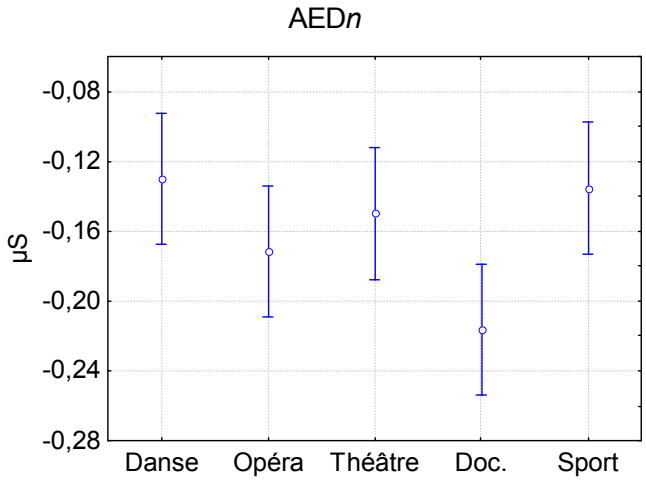


Fig. 8.19. Moyennes d'AED*n* obtenues pour chaque contenu Danse, Opéra, Théâtre, Documentaire (Doc.) et Sport.

Plus précisément, le niveau moyen d'AED*n* était significativement plus faible lors de la visualisation du contenu *Documentaire* par rapport à ceux obtenus pour les contenus *Danse* ($p < 0,05$) et *Sport* ($p < 0,05$). Afin de vérifier si l'augmentation de l'AED au cours de la passation observée dans l'expérimentation A était retrouvée dans cette étude, l'effet de la position (c.-à-d. de l'ordre de présentation du contenu) sur l'AED*n* a été vérifié par une ANOVA (variables indépendantes « Position » et « Participant » -aléatoire-). A l'inverse de la première expérimentation, l'AED n'a pas augmenté au fil du temps mais a, au contraire, montré une diminution significative au cours de la passation (F (4, 100) = 5,01, p <0,01).

8.9.6.4. EFFET DES DEGRADATIONS

Selon les hypothèses formulées (H0p, H1p), un effort mental supplémentaire lié à la présence de dégradations audio et/ou vidéo conduirait à une activation majoritaire du système nerveux sympathique traduite par une augmentation des indices AED, FC et une diminution des indices VSP et TCP. Pour étudier cela, une ANOVA a été réalisée pour chaque contenu, en considérant les variables indépendantes « Période » (P1, P2, P3, P4, éventuellement P5 et A, V, AV, D) et « Participant » (variable aléatoire), et les variables dépendantes « AED*n* », « FC*n* », « VSP*n* », « TCP*n* ». Une forte variabilité inter-individuelle a tout d'abord été constatée avec un effet de la variable « Participant » significatif pour les indicateurs AED*n*, FC*n*, et TCP*n* avec $p < 0,001$. L'ensemble des résultats (effets principaux) est présenté dans l'annexe 8-K_B. Les effets significatifs de la période sont présentés dans le Tableau 8.8 ci-après.

Tableau 8.8. Effets significatifs de la variable indépendante (VI) « Périodes » en considérant la variable aléatoire « Participant » sur les variables dépendantes (VD) « AED*n* », « FC*n*», «TCP*n*» étudiées pour chaque contenu de test Danse, Documentaire, Opéra, Sport et Théâtre.

VI	Contenus	VD	Somme des carrés	ddl effet	ddl erreur	Moyenne des carrés	F	*p*
Période	**Danse**	**AED*n***	1,06	7	175	0,15	23,80	<0,001
	Documentaire	**AED*n***	0,51	8	200	0,06	13,75	<0,001
	Opéra	**AED*n***	0,95	8	200	0,12	21,12	<0,001
		FC*n*	0,02	8	200	0,00	3,38	<0,01
		TCP*n*	0,00	8	200	0,00	2,64	<0,01
	Sport	**AED*n***	0,26	7	175	0,038	9,67	<0,001
		FC*n*	0,02	7	175	0,002	3,83	<0,001
	Théâtre	**AED*n***	0,21	7	175	0,03	7,89	<0,001
		FC*n*	0,01	7	175	0,00	3,68	<0,001
		TCP*n*	0,00	7	175	0,00	3,65	<0,01

La lecture du tableau indique tout d'abord une influence de la période sur l'AED*n* pour chacun des contenus visualisés. Comme l'indique la Figure 8.20 ci-dessous, l'effet trouvé semble être expliqué par le niveau moyen d'AED*n* de la première période de chaque contenu, significativement plus élevé que l'ensemble des périodes suivantes. Comme constaté pour les indicateurs oculaires DP et EBdur, il semble que le début du contenu influence l'activité physiologique des participants.

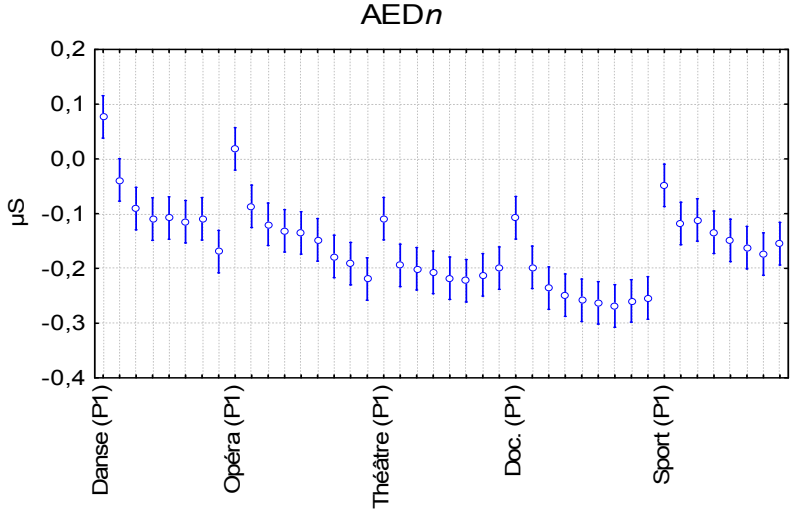


Fig. 8.20. Moyennes d'AED obtenues pour chaque période de chaque contenu visualisé.

La Figure 8.21 ci-dessous, qui présente le signal d'AED*n* obtenu pour un participant donné, constitue une piste d'explication à cet effet. Le niveau plus élevé d'AED*n* au début du contenu pourrait résulter d'une activité résiduelle induite par la phase de complétion de questionnaire à l'origine d'une forte activité physiologique (phase « coûteuse » d'un point de vue cognitif et donc énergétique : effort mnésique, tâche de jugement, lecture/écriture, *etc*.).

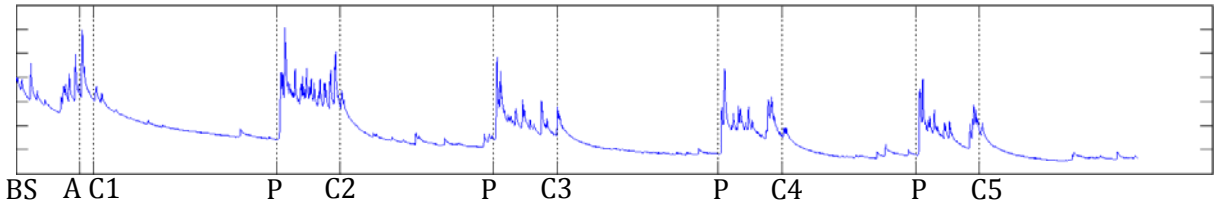


Fig. 8.21. Signal brut d'AED obtenu pour un participant donné tout au long de la passation de test, c'est-à-dire de l'enregistrement de la baseline (BS), à la présentation de l'amorce (A) et des cinq contenus (C1, C2, C3, C4, C5). Le tracé présente également les périodes de pause (P) de 5 min allouées à la complétion des questionnaires.

Un effet de la période a aussi été constaté sur les variations de la FC*n* moyenne pour les contenus *Opéra*, *Sport* et *Théâtre*. Les variations de la FC*n* pour les deux premiers contenus ne permettent pas de conclure à un effet de la période en raison de l'absence de différence significative entre deux séquences adjacentes (voir fig.8.22 ci-dessous). Les variations observées semblent être davantage imputables à une influence liée au contenu.

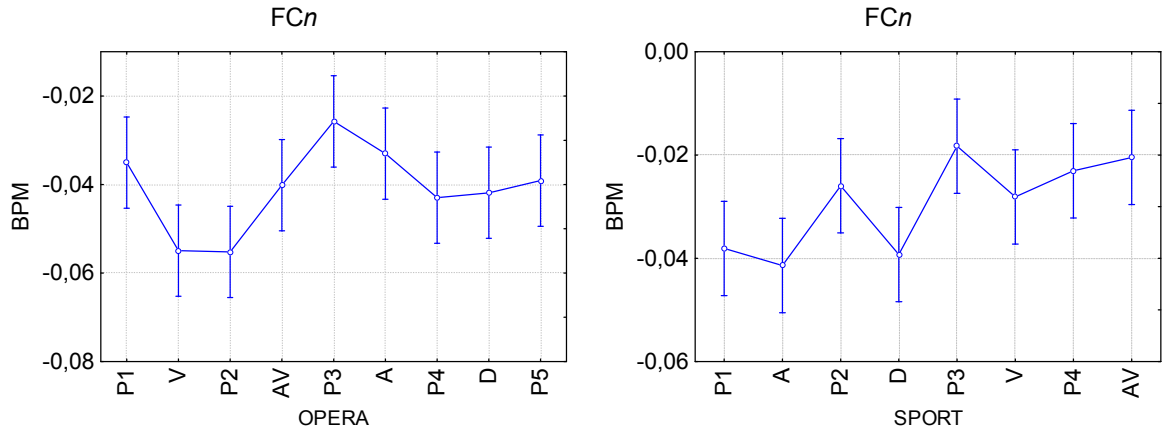


Fig. 8.22. Moyennes obtenues pour chaque période des contenus Opéra et Sport pour l'indice FC*n*. P1, P2, P3, P4 et P5 correspondent aux périodes non dégradées (selon chronologie du contenu). A correspond à la période présentant la dégradation audio, V à la dégradation vidéo, D à la dégradation désynchronisation et AV à la combinaison des dégradations A et V, selon leur ordre d'apparition.

En revanche, les niveaux moyens de FC*n* obtenus pour le contenu *Théâtre* ont montré une augmentation significative entre deux périodes adjacentes à savoir P2 et AV ($p < 0{,}01$), comme le présente la Figure 8-23 ci-après.

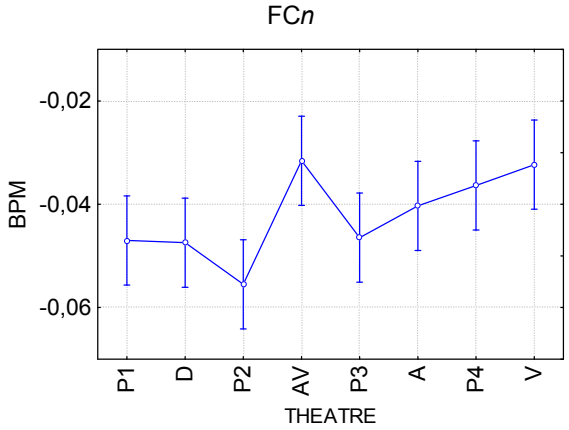


Fig. 8.23. Moyennes obtenues pour chaque période du contenu Théâtre pour l'indice FC*n*. P1, P2, P3 et P4 correspondent aux périodes non dégradées (selon chronologie du contenu). A correspond à la période présentant la dégradation audio, V à la dégradation vidéo, D à la dégradation désynchronisation et AV à la combinaison des dégradations A et V, selon leur ordre d'apparition.

Enfin, l'effet constaté sur la TCP*n* pour les contenus *Opéra* et *Danse* reflétait une tendance linéaire : la température périphérique mesurée durant ces contenus augmentait au fil de la visualisation.

8.9.6.5. EFFET DU TYPE DE NORMALISATION

Deux protocoles de normalisation des données ont été appliqués : une normalisation selon la baseline mesurée au repos et une normalisation à partir d'une baseline mesurée, selon les recommandations de Jennings *et al.* (1992, sect. 3.5, chap. III), durant la réalisation d'une tâche cognitive similaire à la tâche expérimentale mais requérant un niveau d'effort cognitif moins important (amorce). L'objectif était de pouvoir proposer la normalisation la plus adaptée en cas de différences observées. Les analyses à partir des données normalisées par l'amorce ont été réalisées de manière strictement identique à celles effectuées dans les paragraphes précédents. Les résultats n'ont toutefois pas apporté d'information supplémentaire que ce soit pour l'étude des effets de l'activité, du contenu ou des fluctuations de qualité. Ce constat semble indiquer que le choix de l'une ou l'autre méthode de normalisation aboutit à des résultats équivalents pour le type de protocole présenté ici.

8.9.6.6. AUTRES APPROCHES STATISTIQUES

Une des adaptations proposées suite aux résultats de l'expérience A était de varier les approches statistiques toujours pour une étude tonique des signaux recueillis. Comme indiqué dans le chapitre III, différents indicateurs permettent l'étude de la variabilité du rythme cardiaque (VRC) tels que l'étude du SDNN (déviation standard de l'intervalle RR (ms) pour une période donnée, voir § 3.2.6.1, chap. III) ou les indicateurs de basses (BF) et hautes (HF) fréquences issus du domaine fréquentiel et permettant le calcul de ratios conformément aux recommandations de Boonnithi et Phongsuphap (2011, voir § 3.2.6.2, chap. III). Ces derniers préconisent l'étude des indicateurs et ratios suivant : BF normalisée (BFn), différence entre

BF et HF normalisée (dBFHF) et SVI (Sympathovagal Balance Index). Le Tableau 8.9 rappelle le détail du calcul de chaque ratio.

Tableau 8.9. Calcul des ratios BFn, dBFHF et SVI selon Boonnithi et Phongsuphap (2011).

Ratios	Unités	Calcul
BFn	%	100 x BF / (HF + BF+ TBF)
dBFHF	%	| BFn – HFn |
SVI	%	BF/HF

Par ailleurs, l'ensemble des indices physiologiques (FC, AED, TCP et VSP) a aussi été considéré à travers la médiane. Les analyses présentées dans les précédents paragraphes pour l'étude de l'influence du type d'activité (baseline, amorce, C1), du contenu (danse, documentaire, opéra, théâtre, sport) et de la période (P1, P2, P3, P4/ P5 et A, V, AV, D) ont été à nouveau réalisées en considérant les indicateurs de médiane (AED, FC, TCP, VSP) et de VRC (SDNN, BFn, dBFHF, SVI).

Les résultats ont indiqué que globalement la médiane confirmait les effets observés à partir de la moyenne autant pour l'étude de l'influence du type d'activité (Baseline, Amorce, C1), du contenu (Danse, Documentaire, Opéra, Théâtre, Sport) que des dégradations (P1, P2, P3, P4/ P5 et A, V, AV, D) avec notamment un effet de AV sur la FC pour le contenu *Théâtre*.

L'étude du SDNN et des ratios (calculés et moyennés pour chaque vecteur temporel étudié : activité, contenu et périodes) n'a pas permis de dégager d'informations pertinentes.

Enfin, une analyse de corrélations entre mesures subjectives et mesures psychophysiologiques a été conduite. Comme cela pouvait être attendu (les mesures physiologiques et oculaires ayant peu réagi aux conditions de qualité), aucune corrélation pertinente n'a été mise en avant par cette analyse.

8.9.7. CONCLUSIONS MESURES PHYSIOLOGIQUES

L'analyse de **l'effet du type d'activité** (repos, amorce ou visualisation) a montré que le niveau moyen d'AED était plus élevé durant la baseline et l'amorce que lors du premier contenu visualisé. Ce constat pourrait traduire une activation physiologique pendant la baseline et l'amorce en raison d'un **état d'anxiété lié à l'attente de la tâche à venir** (Farah et Sher, 1989, cité par Jennings *et al.*, 1992). L'amorce semble donc avoir été soumise au même biais que la baseline. Ce postulat est conforté par la diminution significative du niveau moyen d'AED observée dès la présentation du premier contenu (mise en route du test). En effet, contrairement à l'augmentation constante observée lors de l'expérimentation A, l'AED a diminué au cours de la passation dans cette dernière étude. Cet effet traduit sans doute la tendance naturelle de l'AED à diminuer au cours du temps en situation de « repos » (sect. 3.3.5, chap. III) qui pourrait ici être associé à un état de relaxation. Ce constat tend à confirmer d'une part, l'état d'irritation/agacement, d'ennui ou de désengagement supposé lors de l'expérimentation A (présentation successive de contenus connus comportant des

dégradations) et d'autre part, qu'**une présentation unique des contenus de test est plus adaptée pour l'étude des mesures physiologiques**.

Le changement d'activité a aussi influencé la fréquence cardiaque, plus faible lors de la présentation de l'amorce que lors de la baseline et de la visualisation du premier contenu. Ce ralentissement pourrait être le **reflet d'une influence des processus attentionnels impliqués dans l'activité de visualisation** d'un contenu audiovisuel (amorce). La différence observée entre l'amorce et l'activité de visualisation pourrait s'expliquer par des différences de contenu, le premier étant simple (du point de vue sémantique : cube en mouvement) et sans dégradations des signaux audio et/ou vidéo contrairement au second. Autrement, la diminution significative de la FC observée durant l'amorce pourrait appuyer l'importance de ce contenu intermédiaire pour préparer le participant aux processus attentionnels engagés dans la tâche de visualisation. Les résultats ne permettent toutefois pas d'étendre son efficacité aux mesures d'AED (reflet d'un état potentiel d'anxiété). L'augmentation linéaire de la TCP entre la baseline et le premier contenu semble difficile à expliquer.

Une **influence du type de contenu** a également été constatée sur l'indicateur d'AED*n* dans la mesure où l'AED*n* était moins élevée durant la visualisation du contenu *Documentaire* par rapport aux contenus *Danse* et *Sport*. Il est intéressant de noter ici que le contenu *Documentaire* correspondait au contenu caractérisé par l'expert par le niveau le plus faible de **dynamique de contenu** (moyennes calculées à partir de l'ensemble des séquences annotées du contenu) tandis que les contenus *Sport* et *Danse* présentaient les niveaux les plus forts. En revanche, la dynamique de caméra était plus faible pour *Danse* que pour *Documentaire*. Ce résultat précise les conclusions émises à la suite de l'expérimentation A et B1 (voir sect. 7.3.1, chap. VII) en indiquant que **l'AED serait sensible aux aspects de dynamique et plus particulièrement à la dynamique de contenu (activité des personnages ou des objets d'intérêt).**

Enfin, l'étude des fluctuations de qualité (dégradations) sur les mesures physiologiques a mis en avant deux principaux effets. Tout d'abord, un **effet marqué du début de contenu,** probablement expliqué par un effet résiduel de l'activité physiologique liée à la complétion de questionnaire, a été observé sur les mesures d'AED*n* qui présentait un niveau moyen systématiquement plus élevé au début de chaque contenu de test. Cet effet a pu être renforcé par le changement d'activité (repos à visualisation) et/ou par un effet d'anticipation lié au début d'un nouveau contenu (toujours averti six secondes avant son commencement). Andreassi, Rapisardi et Whalen (1969) ont constaté que l'AED (tonique) était plus élevée lors de la détection de signaux présentés à intervalles fixes, par rapport à des patterns variables. Ainsi, la présence de l'avertisseur de contenu aurait pu entraîner ou accentuer un phénomène d'anticipation de l'activité de visualisation. Une autre explication pourrait impliquer la présence d'une réponse d'orientation. En effet, Hubert et de Jong-Meyer (1991) ont interprété comme une réponse d'orientation l'augmentation de l'AED constatée durant la première minute de chaque contenu AV de test présenté (10 min, sect. 4.2.2, chap. IV).

Par ailleurs, un **effet de la qualité audiovisuelle** (combinaison des dégradations audio et vidéo) a été observé sur le niveau moyen de FC*n* qui augmentait lorsque celle-ci était

introduite lors du contenu *Théâtre*. Les mesures subjectives peuvent éclairer ce résultat. En effet, les participants ont rapporté avoir éprouvé de la gêne quant à la présentation 3D de ce contenu et une perte de compréhension en présence des dégradations audio et audio-vidéo. Ce même contenu a également été caractérisé par les notes les plus basses de qualité vidéo et audiovisuelle. Ainsi, la combinaison de qualités audio (perte d'intelligibilité, voir sect. 8.9.3) et vidéo (incluant l'altération liée au format 3D) perçues comme gênantes, aurait pu conduire à une activation du SNS (dépenses énergétiques accrues). Cette activation s'expliquerait alors par un effort mental supplémentaire, exprimé par une augmentation de la FC*n*, pour décoder et interpréter le contenu lorsque, à la fois, les signaux audio et la vidéo étaient dégradés. Rappelons ici que dans leurs études Wilson G. M. et Sasse (2000a, 2000b) ont observé que si l'AED était influencée seulement par la dégradation vidéo, la FC quant à elle augmentait en présence d'une dégradation vidéo et aussi en présence d'une dégradation audio lorsque cette dernière correspondait soit à une distorsion soit à une perte de paquets. D'après ce résultat, il semble que la FC soit plus sensible aux dégradations à la fois audio et vidéo. Cette sensibilité permet de penser que, dans le cadre de l'expérimentation C, **l'augmentation constatée de la fréquence cardiaque**[26] **pourrait être expliquée par le cumul des dégradations audio et vidéo (incluant un effet de la 3D).** L'absence d'effet individuel de la dégradation audio ou vidéo sur l'AED et la FC pourrait alors s'expliquer par l'application de dégradations moins fortes dans cette expérience (10% de perte de paquets pour l'audio et réduction du débit vidéo) que celles étudiées par Wilson G. M. et Sasse (20% de perte de paquets pour l'audio et diminution du nombre d'ips). La durée d'application des dégradations était également plus longue dans les études de Wilson G. M. et Sasse (2 min pour l'audio et 5 min pour la vidéo) que dans le cadre de l'expérimentation C (1 min pour l'audio et la vidéo).

L'effet de la qualité sur la FC permet de supposer une modification du pattern de l'activité cardiaque en réaction à un changement de qualité conformément à H1p.

L'étude du protocole de normalisation (repos *vs*. activité) n'a pas permis de privilégier l'une ou l'autre des méthodes utilisées. Enfin, conformément aux perspectives issues de l'expérimentation A, d'autres types d'indicateurs ont été utilisés pour réduire et étudier les données physiologiques. Cependant, ces derniers n'ont pas apporté de nouvelles informations ou d'informations pertinentes dans le cadre de l'étude de l'influence de qualité.

8.10. CONCLUSIONS EXPERIMENTATION C

Les principales conclusions de l'expérimentation C et les explications proposées sont récapitulées dans le Tableau 8.10 ci-dessous.

[26] La différence significative mise avant survenait entre P1 et AV, P1 étant d'une durée de dix secondes. Cette durée peut laisser supposer qu'une méthode adéquate utiliserait une fenêtre d'observation réduite à quelques secondes avant et après l'apparition d'un nouvel évènement. En d'autres termes, une approche phasique pourrait être plus adaptée que la méthode tonique étudiée jusqu'ici.

Tableau 8.10. Récapitulatif des conclusions principales et des interprétations proposées pour chaque type de mesures (Mes.). Les dégradations sont notées de la manière suivante : A pour Audio, V pour vidéo, AV pour audio et vidéo combinée et D pour désynchronisation.

Mes.	Conclusions principales	Explications proposées
SUBJECTIVES	Effet du contenu sur la perception de D (non perçue pour *Danse* et *Sport*)	→Nature non verbale et/ou non diégétique des scènes sonores
	Effet du contenu sur l'évaluation de qualité :	
	- ↓MOSAV et MOV pour *Opéra*	→Effet du débit adaptatif →Effet contaminant de l'expérience hédonique (négative)
	- ↓MOSAV et MOV pour *Théâtre*	→Effet de la 3D
	Effet prédominant de QV sur QAV	→Dépendant du protocole
	Effet de A sur Compréhension et Sentiment d'émotions négatives pour *Théâtre* et *Documentaire*	→Perte d'intelligibilité (lorsque la nature des scènes sonores est essentiellement verbale)
	Effet de AV sur Compréhension et Sentiment d'émotions négatives pour *Théâtre*	→Effet de A
OCULAIRES	Pas d'effet du changement d'activité (entre amorce et 1er contenu visualisé)	-
	Effet du contenu sur DP	→Réflexe photo-moteur →Valence positive/arousal élevé
	Effet du contenu sur SAC	→Dynamique caméra (poursuite oculaire)
	Effet du début du contenu sur DP et EBdur	→Ajustement pupillaire (réflexe photo-moteur) → Phénomènes attentionnels
	Effet du participant	→Forte variabilité inter-individuelle
PHYSIOLOGIQUES	Effet du type d'activité sur AED	→Anxiété, attente
	Effet du type d'activité sur FC	→Engagement attentionnel
	Effet de passation sur AED (diminution)	→Relaxation
	Effet du contenu sur AED	→Dynamique contenu (personnages, objets)
	Effet du début du contenu sur AED	→Effet résiduel de l'activité de complétion →Anticipation →Réponse d'orientation
	Effet de la qualité sur FC (AV)	→Effet du cumul des dégradations (A et V dont 3D)
	Pas d'effet du protocole de normalisation	-
	Pas de nouveaux effets issus de l'analyse complémentaire	-

8.10.1. MESURES SUBJECTIVES

Le questionnaire enrichi a notamment permis d'observer que la dégradation audio influence la *qualité d'expérience* du spectateur en impactant la compréhension du contenu (diminution de l'intelligibilité du message audio). Cet effet n'a pas été reflété par les notes de qualité qui, par conséquent, ne rendent pas compte de l'ensemble des influences des fluctuations de qualité sur la *qualité d'expérience* du spectateur. L'évaluation de l'expérience « hédonique » du spectateur est également intéressante puisque l'attrait (Palhais et *al.,* 2012) ou le rejet (expérimentation C) du contenu pourrait influencer les notes de qualité comme cela est supposé pour le contenu *Opéra* (effet « contaminant »). L'intérêt mais aussi la valence et l'arousal pourraient avoir une influence négative sur les notes de qualité. Il a également été montré que les questions relatives à la détection des dégradations étaient nécessaires notamment pour l'évaluation de la désynchronisation image/son. En revanche, les réponses relatives au niveau de certitude (ce dernier étant fort quand la dégradation est détectée mais mitigé lorsqu'elle ne l'est pas) et au sentiment d'émotions négatives (probablement influencées par l'évaluation de l'impact de la dégradation sur la compréhension) semblent sensibles à un effet du questionnaire. Des études similaires pourraient proposer des questionnaires faisant varier l'ordre de présentation des différents critères évalués pour tester plus amplement la pertinence de ces deux derniers critères. De manière générale, les critères correspondant aux descripteurs de la catégorie Sémantique (modalité, quantité d'information, dynamique) ainsi que l'évaluation de la luminosité pourraient être retirés du questionnaire, les informations qu'ils apportent pouvant être obtenues par une étape en amont de caractérisation des séquences ou des contenus de test. Cela permettrait d'alléger le questionnaire soumis aux participants. Par ailleurs, l'évaluation du niveau général de compréhension pourrait être abandonnée au profit de l'évaluation de l'impact de la dégradation sur le niveau de compréhension plus précis et plus révélateur de l'influence de la qualité sur cet aspect.

Le questionnaire proposé dans cette expérimentation a donc permis d'obtenir un retour plus complet de l'influence de la qualité audio et/ou vidéo sur la *qualité d'expérience* du spectateur. Cependant, l'ensemble des critères évalués n'est pas pertinent. Sur la base des précédentes observations, un questionnaire enrichi pour étudier l'influence de la présence de dégradations audio et/ou vidéo sur la perception de qualité et plus largement sur la *qualité d'expérience* du spectateur est proposé. Le Tableau 8.11 ci-dessous présente le questionnaire suggéré.

Tableau 8.11 : Questionnaire enrichi proposé.

Observable	Statut
QAV, QV, QA	Maintenu
Intérêt	Maintenu
Valence	Maintenu
Arousal	Maintenu
Quantité d'information	caractérisation amont
Compréhension	Abandonné
Modalité	caractérisation amont
Dynamique contenu	caractérisation amont
Luminosité	caractérisation amont
Détection	Maintenu
Certitude de la détection	Abandonné
Impact sur Compréhension	Maintenu
Impact sur Sentiment émotions négatives	A préciser

8.10.2. MESURES PSYCHOPHYSIOLOGIQUES

Un des objectifs de cette expérimentation était de proposer une évolution du protocole utilisé lors de l'expérimentation A tenant compte des effets observés ou supposés des facteurs *passation*, *changement d'activité* (*amorce, avertisseur)*, *habituation, engagement*, *matériel* et *niveau* (durée et seuils de dégradations) sur les mesures psychophysiologiques. De manière générale, les améliorations proposées ont été efficaces, c'est-à-dire qu'elles ont permis d'éviter les biais constatés lors de la première étude (habituation, passation, engagement, matériel). En revanche, l'effet du changement d'activité (de la complétion de questionnaire à la phase de visualisation) n'a pu être évité, le décompte avertissant le participant du commencement de chaque nouveau contenu s'étant révélé insuffisant voir aggravant (phénomène d'anticipation potentiel). Pour éviter cet effet, les mesures enregistrées au cours de la première minute du contenu devraient *a minima* être retirées lors de l'analyse des données.

L'augmentation des seuils et durées de dégradations n'a pas permis l'observation marquée ou systématique d'une influence de la qualité sur le pattern physiologique et oculaire des participants. L'effet de la dégradation audio-vidéo sur la fréquence cardiaque a en effet indiqué que les seuils de dégradations doivent être poussés à l'extrême (cumul d'un effet de la 3D, d'une dégradation vidéo et d'une dégradation audio gênante du point de vue de la compréhension et/ou sentiment d'émotions négatives) pour influencer de manière significative les mesures physiologiques telles qu'étudiées ici (à travers des indicateurs de tendances centrales). Finalement, la 3D a joué un rôle de facteur d'augmentation d'un effort mental seulement lorsque cette dernière introduisait une perte de qualité vidéo.

L'effet observé sur l'activité cardiaque permet néanmoins de croire que les mesures physiologiques peuvent permettre d'observer des fluctuations de qualité, l'augmentation de la fréquence cardiaque pouvant traduire la réalisation d'un effort mental en présence de fortes dégradations. En revanche, aucun effet de fatigue n'a été constaté au niveau de l'activité oculaire probablement parce que les dégradations introduites n'étaient pas suffisantes pour induire un effort mental prolongé ou récurrent (puisque seulement ponctuel) à l'origine d'un état de fatigue.

Il semble que de manière générale les mesures psychophysiologiques recueillies aient principalement été sensibles à des composantes du contenu (luminosité, dynamique) ou à un effet du changement d'activité (pauses et phase de complétion à visualisation). Par ailleurs, une forte variabilité inter-individuelle a été constatée aussi bien pour les mesures oculaires que physiologiques. Cette variabilité aurait pu bruiter les analyses inter-participant effectuées et participer à la non-détection d'un impact de la qualité sur les signaux recueillis. Un effet de la qualité a tout de même été observé, pour un des indicateurs étudié, lorsque des dégradations perçues comme gênantes étaient présentes en même temps sur les modalités audio et vidéo.

Le Tableau 8.12 ci-dessous rappelle les adaptations apportées au protocole de départ et leur statut à l'issue de l'expérimentation C.

Tableau 8.12 : Adaptations du protocole testées dans l'expérimentation C à la fois pour améliorer l'interprétation et la compréhension des mesures subjectives (SUBJ.) et psychophysiologiques. La désynchronisation est indiquée par la lettre « D ».

Mes.	Effets	Adaptation proposée	Statut
SUBJ.	Limite du *questionnaire*	Ajouter une question spécifique à la perception de la désynchronisation	Réalisé
	Effet du facteur *niveau*	Augmenter le seuil de la désynchronisation	Réalisé
	Effet *contenu*	Caractériser les contenus de test	Réalisé
PSYCHOPHYSIOLOGIQUES	Effet du facteur *passation*	Proposer un ordre aléatoire de présentation/ Présenter une seule et unique fois les contenus	Réalisé
	Effet du facteur *activité*	Ajouter un contenu *amorce* Ajouter avertisseur de début de contenu	Réalisé Insuffisant
	Effet du facteur *habituation*	Varier patterns d'introduction des dégradations	Réalisé
	Effet du facteur *engagement*	Favoriser l'engagement sur la tâche	Réalisé
	Effet du facteur *matériel*	Solution de synchronisation	Réalisé
	Effet du facteur *niveau*	Augmenter les durées et seuils des dégradations appliquées	Réalisé Insuffisant
	Effet du facteur *analyse*	Varier les approches	Réalisé Insuffisant
	Effet *contenu*	Caractériser les contenus de test	Perspectives

CONCLUSIONS ET PERSPECTIVES

RAPPEL DES OBJECTIFS

Savoir mesurer l'influence de la qualité audiovisuelle (QAV) restituée sur la *qualité d'expérience* du spectateur (QoE) est aujourd'hui un enjeu majeur dans le domaine de l'offre de services audiovisuels. Les méthodes subjectives actuelles, et notamment la norme UIT-T P.911 (UIT, 1998), limitent cette mesure à la seule évaluation de la qualité perçue du signal audiovisuel restitué. L'objectif du travail présenté dans ce document était d'étendre cette mesure à la mesure de la *qualité d'expérience* du spectateur. Deux types d'indices, subjectifs et psychophysiologiques, ont été proposés pour compléter la mesure de qualité.

L'**ajout de critères subjectifs** à l'évaluation traditionnelle de la qualité audiovisuelle perçue devait permettre d'enrichir l'étude de l'influence de la qualité restituée sur la qualité d'expérience.

Il était également attendu que **des indices psychophysiologiques** complètent la mesure subjective en apportant des informations, non biaisées par l'accès à la conscience (biais des mesures subjectives), relatives au *coût utilisateur* induit par le traitement d'un signal dégradé. Un phénomène de fatigue ou d'effort mental provoqué par la présence de dégradations pourrait en effet réduire la QoE, de façon consciente ou non, et conduire à un rejet du système utilisé.

Un aspect important du travail réalisé a été de proposer un contexte d'évaluation plus proche de conditions réelles de visualisation. Ce choix était motivé par l'ambition d'obtenir un retour plus représentatif de l'impact de la qualité audiovisuelle, sur la qualité d'expérience du spectateur, que ne le permettent les séquences de 10 s recommandées par les normes actuelles. Pour cela, des contenus de plusieurs minutes, plus adéquates pour étudier les différentes influences de la qualité sur le spectateur (notamment en matière de fatigue ou d'effort), ont été présentés avec des niveaux fluctuants de qualité (c.-à-d. avec des dégradations qui n'étaient pas toujours équivalentes ou de même nature au sein d'un contenu donné).

La méthode proposée a été élaborée sur la base des travaux de Wilson G. M. et Sasse (2000a, 2000b) portant sur l'ajout de mesures toniques physiologiques aux méthodes subjectives actuellement utilisées pour évaluer la qualité audio et/ou vidéo. Il était attendu que la présence de dégradations durant une activité de visualisation de contenus AV 2D ou 3D diminue la *qualité d'expérience* envisagée non plus sous le seul angle de la qualité perçue mais aussi du point de vue d'autres aspects déterminants pour la QoE, évalués à partir de mesures subjectives additives (questionnaire enrichi) et psychophysiologiques.

LE CAS DES MESURES PHYSIOLOGIQUES ET OCULAIRES

APPORTS

L'ajout des mesures physiologiques et oculaires répondait à deux principaux besoins **(1)** échapper aux biais des mesures subjectives **(2)** considérer l'étude de l'influence de la QAV sur la *qualité d'expérience* du point de vue du *coût utilisateur* (effort mental, fatigue). Globalement, les mesures psychophysiologiques n'ont pas permis d'atteindre ces objectifs. Il a en effet été assez difficile d'observer des effets de la qualité sur les indicateurs étudiés. Par conséquent, la question de la relation entre mesures subjectives et mesures psychophysiologiques (par exemple le lien entre la présence d'un effort mental ou de fatigue sur l'expérience subjective) n'a pu être entièrement traitée par ces travaux.

L'activité oculaire n'a pas montré de réponses aux fluctuations de qualité ; en d'autres termes, elle n'a pas reflété l'apparition, supposée séquentielle, d'effort mental et de fatigue. Ces mesures ont en revanche été affectées par le contenu, notamment le niveau de luminosité, et par le changement d'activité (repos *vs*. visualisation et complétion de questionnaire *vs*. reprise de la visualisation). Ces mêmes effets ont été observés pour les signaux physiologiques.

Néanmoins, ces derniers ont montré une meilleure sensibilité, quoiqu'assez faible, aux dégradations présentées. Plus précisément, l'activité électrodermale augmentait lorsqu'un effet des dégradations audio et/ou vidéo se cumulait à un effet d'irritation/agacement, *etc.* lié à la présentation successive d'un même contenu, la plupart du temps dégradé (chaque contenu était vu et entendu trois fois de suite). L'activité électrodermale est reconnue pour sa spécificité à refléter l'arousal. Ainsi, la réponse observée pourrait traduire une augmentation de l'arousal (activation physiologique) en raison de l'état d'irritation provoqué par la présentation répétitive de contenu cumulée à la présence de dégradations. De la même manière, la perte de qualité a été exprimée sur le plan physiologique par une augmentation de la fréquence cardiaque lorsqu'un effet de la dégradation audio (diminuant l'intelligibilité du contenu) et de la dégradation vidéo (incluant un effet de la dégradation vidéo et un effet dégradant de la 3D), perçues comme gênantes, s'additionnait. L'effet sur la fréquence cardiaque tend à corroborer les résultats de Wilson G. M. et Sasse quant à la plus grande sensibilité de cet indice, par rapport à celle de l'activité électrodermale, aux dégradations à la fois vidéo et audio. Dans ce document, l'augmentation de la FC pourrait traduire un effort mental lorsque les conditions de visualisation sont fortement dégradées. Ces observations conduisent à la conclusion suivante : **l'effet de la qualité sur l'activité physiologique est observé lorsque des dégradations gênantes se cumulent et/ou que les conditions de passation sont pénibles pour le participant**.

Cependant, les effets constatés de la qualité se limitaient à la réponse d'un seul indicateur physiologique à la fois. Ce **manque de redondance** peut être mis en regard du concept d'activation multidimensionnelle supposant qu'une activité donnée correspondrait à un pattern d'activation spécifique pour lequel les performances sont optimales (sect. 4.2.2, chap. IV). Les signaux physiologiques pourraient alors répondre plus spécifiquement à certaines

dégradations ou à certains contextes de dégradations. Toutefois, un effet contextuel peut aussi être envisagé. Ce dernier point permet difficilement de conclure à un réel impact de la qualité sur les indicateurs physiologiques. **Dans tous les cas, les indicateurs psychophysiologiques sont peu propices à l'étude d'effets subtils de la qualité ou d'effets qui ne seraient pas déjà reflétés par les mesures subjectives.**

Par ailleurs, un des intérêts des mesures physiologiques est de pouvoir refléter l'activité de l'organisme en temps réel, offrant ainsi la possibilité d'étudier l'influence des dégradations au moment où elles surviennent. Cette spécificité a présenté l'avantage de pouvoir proposer un contexte de visualisation plus réaliste (contenus d'une durée > 10 s et qualité fluctuante), plus proche d'une évaluation *in situ*. Cependant, le port des capteurs physiologiques, la présence des caméras de l'*eye tracker* et la perte de mobilité liée à l'enregistrement des mesures limitent la représentativité du contexte proposé. Ces contraintes n'ont pour autant pas été exprimées du point de vue psychophysiologique (pas de fatigue). Elles pourraient par contre avoir participé, au moins en partie, à l'augmentation du niveau de fatigue rapporté par les participants après le test.

D'un point de vue méthodologique, les mesures physiologiques et oculaires ne se sont pas révélées pertinentes pour pouvoir étudier la *qualité d'expérience*, du point de vue du *coût utilisateur*, dans le cadre d'un contexte proche de conditions réelles de visualisation. Ces mesures n'ont pas permis de pallier les biais des mesures subjectives en apportant des informations complémentaires pour l'étude de la *qualité d'expérience* du spectateur.

POUR ALLER PLUS LOIN…

Différents postulats ont été émis pour expliquer le peu d'influence de la qualité sur les mesures psychophysiologiques. Malgré les précautions prises, un certain nombre de facteurs rendent difficile l'interprétation de ces mesures et peuvent avoir desservi l'observation d'un effet de la qualité, tels que :

- la **passation :** agacement/irritation ou au contraire relaxation/repos,
- le **changement d'activité** : repos *vs*. visualisation, complétion questionnaire *vs*. reprise de visualisation (effet résiduel, engagement attentionnel, réflexe photo-moteur, anticipation, *etc*.),
- le **contenu** : influence particulièrement les mesures. La caractérisation réalisée a permis de mieux comprendre les paramètres les plus « impactant ». La dynamique (contenu) et la luminosité influencent particulièrement les mesures. Si la caractérisation offre une meilleure compréhension des variations observées, elle ne permet pas de s'affranchir de cette influence. Des contenus de « laboratoire » (dont l'ensemble des paramètres du contenu est connu et maitrisé) pourraient offrir un meilleur contrôle de ces effets. Cependant, leur réalisation supposerait un coût élevé de compétences, de matériel et de temps. Au-delà de ces aspects, de tels

contenus risqueraient surtout de poser le problème de la représentativité des séquences élaborées,

- l'**état émotionnel** induit par le contenu : rejet ou adhésion,
- le **protocole** : habituation, niveaux et durées des dégradations, anticipation.

Par ailleurs, seule l'activité physiologique tonique a été étudiée sur la base de comparaison de moyennes entre différentes fenêtres temporelles correspondant à des conditions expérimentales données (dégradées *vs*. non dégradées par exemple). Cette approche statistique suppose la suppression d'un grand nombre d'informations présentes sur le signal, pouvant avoir masqué certains effets. **L'étude de l'activité phasique** par exemple à partir des réponses électrodermales (RED : voir annexe 3-A pour plus de précisions) pourrait être une approche plus adaptée. Pour explorer cette piste, des analyses supplémentaires ont été conduites dans le cadre de ce travail. Une première étape a consisté à extraire des indicateurs de l'activité phasique (RED, variations de l'enveloppe d'amplitude du volume sanguin périphérique, différentiel de la température cutanée par pas de 5 s, *etc.*). Dans un second temps, les signaux recueillis lors de l'expérimentation C (activité électrodermale, diamètre pupillaire, *etc*.) et les indicateurs phasiques ont été moyennés à la seconde. Plusieurs de ces indicateurs ont ensuite été analysés au sein d'un même modèle de Markov caché. L'idée était de proposer un modèle empirique de détection automatique de l'influence de dégradations sur le pattern physiologique de l'individu par combinaison d'indicateurs (voir Lassalle, Daniel, Le page, Goujon et Gros, 2013, papier présenté en annexe 9-A). Cette approche n'a toutefois pas permis d'observer de réactions systématiques en présence de dégradations. En revanche, le contenu semble expliquer un certain nombre d'évènements physiologiques. La qualité de la détection effectuée doit être améliorée en envisageant par exemple un meilleur apprentissage et en tenant compte des différents facteurs d'influences de ces mesures (contenu, protocole, individu).

La forte **variabilité inter-individuelle** constatée autant pour les mesures oculaires que physiologiques peut également avoir participé à la difficulté d'observer des réponses psychophysiologiques aux variations de qualité audiovisuelle. Comme discuté précédemment, les signaux recueillis était analysés à partir de moyennes, celles-ci regroupaient alors des individus pouvant avoir des patterns de réponses opposés (par exemple certains individus pourraient être moins sensibles aux dégradations). La moyenne est directement influencée par ces différences la rendant peut-être peu adaptée à l'analyse des mesures psychophysiologiques. La constitution de groupes distincts d'individus pourrait alors être envisagée. La littérature fait état de deux types de profils : les profils labiles qui montrent une plus grande réactivité à la fois électrodermale et cardiaque et sont plus résistants au phénomène d'habituation et les profils stabiles qui présentent une habituation rapide (Dawson *et al.*, 2007, p.173). Ce type de profil est stable dans le temps et pourrait donc être déterminé pour un individu donné (les individus labiles présentant un taux élevé de RED-NS au repos tandis que les profils stabiles montrent peu de RED-NS, Dawson *et al.*, 2007). Une approche par *clustering* pourrait également permettre l'identification de différents groupes de patterns de réactions physiologiques (en fonction du sexe, du profil labiles *vs*. stabiles, *etc.*).

Enfin, **la présence seule de dégradations n'est peut-être tout simplement pas un élément suffisant** pour entraîner un surcroît de dépenses énergétiques (effort mental puis fatigue) comme tend à le montrer l'effet de la qualité audiovisuelle sur l'activité physiologique constaté seulement lorsque des dégradations et/ou des conditions de passation pénibles pour le participant se cumulent. Les résultats de Wilson G. M. et Sasse ne vont pas dans le sens de cette supposition. Cependant, ces auteurs ont proposé des dégradations appliquées sur de plus longues périodes : les dégradations audio étaient présentées durant deux minutes lors de séquences uniquement verbales (perte d'intelligibilité probable) et les dégradations vidéo étaient introduites durant cinq minutes lors de séquences audiovisuelles. Les effets les plus forts de la qualité sur l'activité physiologique, rapportés par ces auteurs, ont été observés en réponse à la dégradation vidéo (diminution du nombre d'image par seconde). Il se peut que la nature de cette dégradation soit plus gênante pour l'individu que les dégradations étudiées dans ce document. De plus, une tâche était demandée aux participants durant la visualisation des séquences audiovisuelles (jugement des candidats lors d'un entretien et évaluation de la qualité audiovisuelle), contexte sans doute plus favorable au maintien de l'engagement attentionnel du participant durant toute la visualisation. Dans l'idée de proposer un contexte capable de mettre en évidence un effort mental en présence de dégradations, une piste intéressante pourrait impliquer davantage le participant par l'introduction d'une tâche explicite à réaliser durant l'activité de visualisation (dénombrement d'objets spécifiques présents dans une scène par exemple). La *qualité d'expérience* serait alors étudiée à travers des mesures de performances, des mesures subjectives et des mesures de l'activité physiologique et oculaire de l'individu, comme le suggère Wastell et Newman (1996). Une expérimentation annexe (voir annexe 9-B) a étudié la pertinence d'un tel protocole. Les résultats indiquent que les effets de la qualité audiovisuelle sur les mesures physiologiques et oculaires sont, à nouveau, peu marqués. Toutefois, un effet de la dégradation (selon le type de tâche et de séquence) sur les mesures psychophysiologiques peut être supposé. Ce constat permet de croire que le protocole proposé est une piste prometteuse pour l'étude de l'influence de la qualité audiovisuelle à travers ce type de mesures. Plus généralement, un protocole de « double-tâche » (visualisation et tâche explicite) pourrait être l'approche appropriée pour étudier l'influence de la qualité audiovisuelle, notamment concernant le coût pour le spectateur, à partir de mesures de performances, de mesures subjectives et des mesures psychophysiologiques.

Les différentes influences pour expliquer la difficulté à observer un effet de la qualité audiovisuelle sur les indicateurs physiologiques et oculaires sont illustrées dans le schéma (fig. 9.1) ci-dessous.

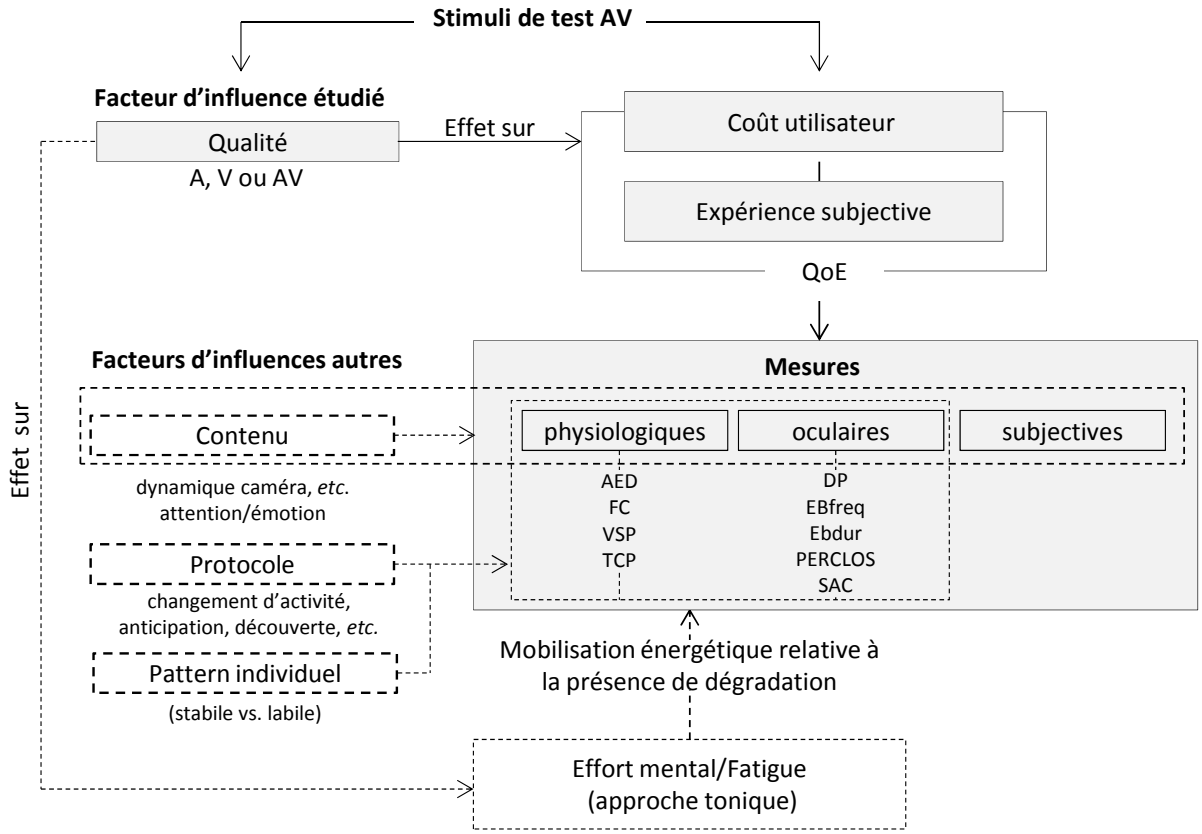


Fig. 9.1. Schéma présentant les différents facteurs d'influences sur les mesures étudiées dans le cadre de l'approche hybride proposée.

Les différentes précautions destinées à s'affranchir d'un certain nombre de biais propres aux mesures psychophysiologiques (caractérisation du contenu), les améliorations apportées au protocole (plus grande implication du participant, solution de synchronisation, présentation unique des contenus, augmentation des seuils et durées de dégradations, introduction d'un contenu *amorce*, ajout d'une tâche) ou la diversification des analyses conduites (type de normalisation des données, analyses de différents indicateurs toniques puis phasiques, *etc.*) n'ont pas suffit à observer un effet franc de la qualité sur les réponses physiologiques et oculaires étudiées. Ces mesures sont extrêmement sensibles à un grand nombre de facteurs souvent difficile à maitriser. En conséquence, il paraît difficile d'isoler des effets propres aux facteurs expérimentaux étudiés tout au moins dans le cadre de l'évaluation de qualité. L'étude des mesures psychophysiologiques implique de plus un coût élevé en matière de configuration logicielle, de traitement et d'analyses. Les attentes formulées au départ, quant à la méthode hybride proposée, pour pallier les faiblesses des méthodes actuelles doivent être réactualisées en faveur des mesures subjectives.

LE CAS DES MESURES SUBJECTIVES

APPORTS

A l'inverse des mesures psychophysiologiques, les mesures subjectives ont été sensibles à la qualité audiovisuelle non seulement sur le plan de la qualité perçue mais aussi de la *qualité d'expérience* étudiée à l'aide de critères subjectifs additifs. Un questionnaire enrichi a pu être proposé à l'issue des expérimentations réalisées.

Concernant l'évaluation de qualité, un apport important consiste à **évaluer séparément les qualités audio et vidéo**. La norme UIT-T P.911 ne recommande actuellement que l'évaluation de la qualité audiovisuelle seule. Or les expérimentations présentées montrent que les notes de qualité audio et vidéo précisent la note de qualité globale en indiquant la modalité ayant le plus fortement participé à son élaboration. Il semble d'ailleurs que la participation prédominante de la qualité d'une modalité sur l'autre, à l'élaboration de la note de qualité audiovisuelle, dépend du contenu mais aussi plus largement du protocole testé (influence probable de la durée de la séquence, de la durée et du seuil des dégradations, répétition des contenus, *etc.*). Les influences mutuelles entre qualité audio et vidéo ainsi que leurs contributions individuelles à l'évaluation de QAV pourraient être étudiées plus spécifiquement en fonction des différents protocoles testés.

L'évaluation séparée des qualités audio et vidéo a aussi permis de constater que les participants intègrent les problèmes de désynchronisation dans la note de qualité audio et non dans celle de qualité audiovisuelle, pouvant rendre difficile l'interprétation des notes de qualité audio et vidéo. Ce constat montre qu'il est judicieux d'ajouter au questionnaire d'évaluation de qualité une question spécifique à la détection de désynchronisation.

Le questionnaire enrichi a également amélioré l'étude de l'influence de la qualité restituée sur la *qualité d'expérience* du spectateur. Un résultat principal concerne la limite des notes de qualité à refléter certains effets des dégradations. **La dégradation audio a notamment occasionné une perte de compréhension et un sentiment d'émotions négatives** diminuant par conséquent la *qualité d'expérience*. Cet effet n'a pas été reflété par les notes de qualité audio. L'influence de cette dégradation s'explique probablement par une perte d'intelligibilité lorsqu'elle est introduite au cours de contenus principalement « verbaux diégétiques ». Ce constat tend à préciser ceux de la « Commission 12 » de l'UIT (COM12-61-E, 1998) supposant que dans un contexte conversationnel interactif, la vidéo serait jugée à partir de critères de qualité (netteté, fluidité) tandis que l'audio serait évalué sur la base de critères d'acceptabilité (intelligibilité, volume, écho). Les résultats obtenus dans le cadre de cette thèse tendent à démontrer que cela est également vrai pour un contexte passif si les contenus présentés sont verbaux. Ainsi, **pour un contexte passif**, **les participants jugeraient la qualité audio et vidéo lorsque le contenu audio est majoritairement non verbal et/ou non diégétique et la qualité vidéo et l'acceptabilité audio lorsque le contenu audio est majoritairement verbal et diégétique**. Ce résultat montre bien que les échelles de qualité ne sont pas suffisantes pour rendre compte fidèlement de l'effet de la qualité audio lorsque celle-

ci altère l'intelligibilité et plus généralement de l'effet de la qualité restituée sur la *qualité d'expérience* du spectateur. Ce dernier point attire l'attention sur les limites de l'étude de l'influence de la qualité sur le spectateur à travers les échelles recommandées par l'UIT et la nécessité de proposer un questionnaire plus étendu.

Il est aujourd'hui largement reconnu que la perception de qualité est dépendante du contenu. De ce fait, la sélection des séquences de test est une étape critique. La **caractérisation des séquences de test** réalisée dans le cadre de ce travail autorise plusieurs conclusions. Premièrement, la dégradation de désynchronisation doit, pour être perçue, survenir sur des contenus présentant des scènes audio à la fois verbales et diégétiques (son en relation avec l'image). Dans le cas contraire, son effet ne pourra être étudié, les notes obtenues, le cas échéant, ne pourront donc pas se prétendre représentatives de cette dégradation.

La caractérisation a aussi permis de mieux comprendre le lien entre contenu et perception de qualité. Comme cela pouvait être attendu, une dégradation audio diminue davantage le niveau perçu de QAV lorsque la nature de la modalité dominante est audio. Le même constat peut être apporté pour une dégradation vidéo appliquée lors de séquences caractérisées par une modalité dominante vidéo. Il a également été constaté, dans le cadre des conditions de dégradations appliquées, que :

(a) une rupture du flux vidéo (saccades) sur des séquences de dynamique forte (modalité dominante vidéo) dégraderait plus fortement la qualité perçue que la perte de netteté (diminution du débit),
(b) la perte de netteté vidéo sur des séquences de dynamique faible (et de modalité dominante vidéo) dégraderait plus fortement la qualité perçue qu'une rupture du flux vidéo,
(c) une rupture du flux audio (perte de paquets) dégraderait plus fortement la qualité perçue qu'une perte de netteté (débit).

La rupture de continuité du flux audio entraînerait une diminution de l'intelligibilité tandis que la perte de netteté vidéo constituerait une perte d'information visuelle plus importante que la rupture de continuité.

En-dehors du lien entre contenu et perception de qualité, la QoE est aussi influencée par le contenu en tant que tel. Dans les études réalisées, il semblerait que les **séquences faiblement dynamiques**, *a fortiori* audio (pour le corpus utilisé), **ont entraîné une expérience hédonique plutôt négative** tandis que des **séquences dynamiques** et *a fortiori* vidéo **ont entraîné une expérience hédonique plutôt positive**. Par ailleurs, l'expérience hédonique suscitée par le contenu tend à avoir un effet « contaminant » sur l'évaluation de qualité vidéo et audiovisuelle : lorsque le contenu est peu apprécié, les notes de qualité diminueraient (expérience C).

Cette dépendance au contenu insiste sur **l'importance de l'étape de sélection des séquences ou contenus de test et de leur caractérisation.** La norme UIT-T P.911 propose déjà une caractérisation des contenus audio et vidéo, considérés séparément. **Il conviendrait d'ajouter**, à cette première description, **les descripteurs du rapport entre audio et vidéo (diégétique), de modalité dominante et/ou de dynamique** (ces deux critères étant fortement liés dans les études présentées) en raison de leurs influences sur la perception de qualité

D'un point de vue méthodologique, un ensemble de recommandations peut être apporté pour l'évaluation de l'influence de la qualité audiovisuelle restituée sur la *qualité d'expérience*, à savoir :

- la **nécessité de réaliser une caractérisation des séquences de test** notamment en matière de modalité dominante, de dynamique et de diégétique. La nature verbale ou non verbale de la séquence doit également être définie,
- **l'évaluation séparée des qualités audio et vidéo**,
- **l'ajout de questions spécifiques à la détection des dégradations** et notamment de désynchronisation, lorsque celle-ci est étudiée. Dans le cas contraire, les participants attribuent la dégradation à la modalité audio, n'ayant pas la possibilité de juger la désynchronisation autrement qu'à partir des échelles dont il dispose,
- **l'utilisation d'un questionnaire étendu** de l'évaluation de la qualité à l'évaluation de la *qualité d'expérience* du spectateur (hédonique, compréhension, détection, émotions).

POUR ALLER PLUS LOIN…

Il pourrait également être intéressant d'affiner le questionnaire réalisé par l'ajout de critères supplémentaires pour l'étude de la QoE tels que l'effort mental ressenti, l'état émotionnel du spectateur (frustration, tension, énervement, *etc.*), l'agréabilité de la séquence, *etc.* L'étude annexe citée précédemment a permis d'initier un travail en ce sens (annexe 9-B). Son objectif était d'étudier le *coût utilisateur* lors de dégradations selon l'approche de Wastell et Newman (1996), c'est-à-dire à partir de l'analyse conjointe de performances, de mesures subjectives et de mesures psychophysiologiques. Les résultats obtenus indiquent une diminution des performances ainsi qu'une augmentation du sentiment de frustration et de l'effort mental ressenti lorsque les tâches sont réalisées en présence de dégradations. De manière générale, les séquences dégradées sont jugées comme étant moins intéressantes, moins compréhensibles, moins stimulantes, moins originales et moins agréables. La sensibilité, à la présence de dégradations, des indicateurs subjectifs testés est une piste encourageante pour l'élaboration d'un questionnaire plus pertinent pour l'étude de l'influence de la qualité restituée sur la *qualité d'expérience* du spectateur.

Les questionnaires testés n'ont toutefois pas exploré les aspects de criticité des dégradations introduites. En effet, malgré la diminution du niveau perçu de qualité, de compréhension et le sentiment d'émotions négatives (comme la frustration, l'énervement,

provoqués par la visualisation d'un contenu dégradé, il se peut que la *qualité d'expérience* demeure acceptable (au sens du rejet ou non du système) pour le spectateur. Le seuil d'acceptabilité relatif non pas à la qualité perçue mais bien à la *qualité d'expérience* pourrait alors être plus spécifiquement étudié. Par exemple, un critère pourrait être la possibilité pour le participant de stopper l'activité de visualisation.

Enfin, une critique d'ordre général peut être émise concernant les échelles, principalement à trois niveaux, utilisées pour évaluer les critères subjectifs additifs testés. Le nombre d'items disponible aurait pu limiter la précision de l'évaluation du participant. De manière générale, la majorité des critiques faites aux normes actuelles d'évaluation et relatives aux biais des mesures subjectives sont ici retrouvées. Ce dernier point souligne encore une fois l'importance de rechercher des indicateurs complémentaires, par exemple la mesure de performances, pour optimiser les méthodes d'évaluation.

Le contexte actuel de l'offre de services audiovisuels est extrêmement compétitif. Dans l'objectif d'offrir des services concurrentiels toujours plus innovants mais dont la *qualité d'expérience* demeure optimale pour l'utilisateur, l'évaluation de la QoE est un élément clé. Les notes de qualité doivent clairement être complétées afin de mieux comprendre la manière dont la qualité restituée influence la QoE. Le questionnaire suggéré à l'issue des différentes expérimentations est une base pertinente pour une mesure plus représentative de la *qualité d'expérience* du spectateur. Par ailleurs, l'étape de caractérisation des contenus et séquences de test est un préalable indispensable pour pouvoir expliquer et comprendre la manière dont la qualité des signaux audio et vidéo conditionne la qualité d'expérience finale.

L'apport méthodologique majeur de ce travail consiste donc en une contribution pour une méthode d'évaluation de la qualité plus complète que celles actuellement recommandées par l'Union Internationale des Télécommunications, c'est-à-dire, une méthode ne portant non plus sur la seule évaluation de la perception de qualité des signaux audio et/ou vidéo mais sur la *qualité d'expérience* du spectateur dans son ensemble. Ce travail a mis l'accent sur la capacité des mesures subjectives à apporter des informations pertinentes sur la *qualité d'expérience* du spectateur.

REFERENCES

Ahlstrom, U. et Friedman-Berg, F. J. (2006). Using eye movement activity as a correlate of cognitive workload. *International Journal of Industrial Ergonomics, 36*(7), 623-636.

Akselrod, S., Gordon, D., Ubel, F. A., Shannon, D. C., Berger, A. et Cohen, R. J. (1981). Power spectrum analysis of heart rate fluctuation: a quantitative probe of beat-to-beat cardiovascular control. *Science, 213*(4504), 220-222.

Allport, D. A. (1980). Patterns and actions: Cognitive mechanisms are content-specific. In G. Claxton (Ed.), *Cognitive psychology: New directions* (p. 32-59). London, R. –U.: Routledge & Kegan Paul.

Amiar., Y. (1995). L'analyse du film et de l'image fixe: approche méthodologique *Revue Recherche sur l'information scientifique et technique, 5*(2), 23-28.

Andreassi, J. L. (2007). *Psychophysiology: Human behavior and physiological response* (5e éd.). Mahwah, New Jersey : Lawrence Erlbaum.

Andreassi, J.L., Rapisardi, S.C. et Whalen, P.M. (1969). Autonomic responsivity and reaction time under fixed and variable signal schedules. *Psychophysiology*, *6*, 58–69.

Appel, M. L., Berger, R. D., Saul, J. P., Smith, J. M. et Cohen, R. J. (1989). Beat to beat variability in cardiovascular variables: noise or music? *Journal of the American College of Cardiology, 14*(5), 1139-1148.

Arrighi, R., Alais, D. et Burr, D. (2006). Perceptual synchrony of audiovisual streams for natural and artificial motion sequences. *Journal of Vision, 6*(3), 260-268.

Backs, R. W. et Boucsein, W. (2000). *Engineering psychophysiology: issues and applications*. Mahwah, New Jersey: Lawrence Erlbaum Mahwah.

Baddeley, A. D. (1986). *Working memory*. Oxford, R. –U. : Clarendon.

Bahill, A. T. et Stark, L. (1975). Overlapping saccades and glissades are produced by fatigue in the saccadic eye movement system. *Experimental Neurology, 48*(1), 95-106.

Bahrick, L. E. (1987). Infants' intermodal perception of two levels of temporal structure in natural events. *Infant Behavior and Development, 10*(4), 387-416.

Barry, R. J., Clarke, A. R., McCarthy, R., Selikowitz, M. et Rushby, J. A. (2005). Arousal and activation in a continuous performance task: An exploration of state effects in normal children. *Journal of psychophysiology, 19*(2), 91-99.

Barry, R. J. et Sokolov, E. N. (1993). Habituation of phasic and tonic components of the orienting reflex. *International Journal of Psychophysiology, 15*(1), 39-42.

Bauer, L. O., Strock, B. D., Goldstein, R., Stern, J. A. et Walrath, L. C. (1985). Auditory discrimination and the eyeblink. *Psychophysiology, 22*(6), 636-641.

Baumstimler, Y. et Parrot, J. (1971). Stimulus generalization and spontaneous blinking in man involved in a voluntary activity. *Journal of experimental psychology, 88*, 95-102.

Beatty, J. et Lucero-Wagoner, B. (2000). The pupillary system. Dans J. T. Cacioppo, L. G. Tassinary et Berntson, G.G. (dir.), *Handbook of psychophysiology* (p.142-162). New York, NY: Cambridge University Press.

Bechara, A., Damasio, H., Damasio, A. R. et Lee, G. P. (1999). Different contributions of the human amygdala and ventromedial prefrontal cortex to decision-making. *The Journal of Neuroscience, 19*(13), 5473-5481.

Beerends, J. et De Caluwe, F. (1999). The influence of video quality on perceived audio quality and vice versa. *Journal of the Audio Engineering Society (JAES), 47*(5), 355-362.

Bermant, R. I. et Welch, R. B. (1976). Effect of degree of separation of visual-auditory stimulus and eye position upon spatial interaction of vision and audition. *Perceptual and motor skills, 43*(2), 487-493.

Berntson, G. G., Cacioppo, J. T. et Fieldstone, A. (1996). Illusions, arithmetic, and the bidirectional modulation of vagal control of the heart. *Biological Psychology, 44*(1), 1-17.

Bertelson, P. et Radeau, M. (1981). Cross-modal bias and perceptual fusion with auditory-visual spatial discordance. *Attention, Perception et Psychophysics, 29*(6), 578-584.

Berthoz, A. et Petit, L. (1996). Les mouvements du regard: une affaire de saccades. *La Recherche, 289*, 58-65.

Boonnithi, S. et Phongsuphap, S. (2011). Comparison of heart rate variability measures for mental stress detection. Dans *Proceedings of the IEEE Computing in Cardiology conference*, *38*, 85-88.

Boucsein, W. (1993). Psychophysiology in the workplace—goals and methods. *International Journal of Psychophysiology, 14*(2), 115.

Boucsein, W. (2012). *Electrodermal activity* (2e éd.). New York, NY: Springer.

Bradley, M. M. et Lang, P. J. (1994). Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry, 25*(1), 49-59.

Bradley, M. M., Miccoli, L., Escrig, M. A. et Lang, P. J. (2008). The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology, 45*(4), 602-607.

Cacioppo, J. T., Berntson, G. G., Larsen, J. T., Poehlmann, K. M. et Ito, T. A. (2000). The psychophysiology of emotion (2e éd.). Dans M. Lewis et J.M. Haviland –Jones (dir.), *Handbook of emotions* (p. 173–191). New York, NY: Guilford.

Cacioppo, J. T. et Tassinary, L. G. (1990). Psychophysiology and psychophysiological inference. Dans J. T. Cacioppo et L. G. Tassinary (dir.), *Principles of psychophysiology: Physical, social, and inferential elements* (p. 3-33). New York, NY: Cambridge University Press.

Cain, B. (2007). *A review of the mental workload literature* (rapport OTAN N° RTOTR-HFM-121-Part-II). Toronto, Canada : Defense Research and Development. Récupéré du site http://www.dtic.mil/dtic/tr/fulltext/u2/a474193.pdf

Calcanis, C., Callaghan, V., Gardner, M. et Walker, M. (2008). Towards end-user physiological profiling for video recommendation engines. Dans *Proceedings of the 4th International Conference on the Intelligent Environments*, 1-5.

Campbell, R.. et Dodd, B. (1980). Hearing by eye. *Quarterly Journal of Experimental Psychology, 32*(1), 85-99.

Cannon, W. B. (1915). *Bodily changes in pain, hunger, fear and rage. An account of recent researches into the function of emotional excitement*. New York, NY : D. Appleton & Co.

Cannon, W. B. (1927). The James-Lange theory of emotions: A critical examination and an alternative theory. *The American Journal of Psychology, 39*(1/4), 106-124.

Cavé, C., Ragot, R. et Fano, M. (1992). Perception of sound-image synchrony in cinematographic conditions. Dans *Proceedings of the fourth workshop on rhythm perception and production*, 25-35.

Chanquoy, L., Tricot, A. et Sweller, J. (2007). *La charge cognitive*. Paris, France : Armand Colin.

Chapman, L. J., Chapman, J. P. et Brelje, T. (1969). Influence of the experimenter on pupillary dilation to sexually provocative pictures. *Journal of Abnormal Psychology, 74*(3), 396-400.

Chateau, N. (1997). Perceptual fusion of audiovisual virtual stimuli. Dans *Proceedings of TNO-TM report*, Soesterberg, Pays-Bas.

Chen, W. (2012). *Caractérisation multidimensionnelle de la qualité d'expérience en télévision de la TV3D stéréoscopique* (thèse de Doctorat, Université de Nantes Angers, Le Mans). Récupéré du site thèse en ligne : http://tel.archives-ouvertes.fr/docs/00/78/59/87/PDF/ThA_se_final_Chenwei.pdf

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *The Journal of the Acoustical Society of America, 25*(5), 975-979.

Clarion, A. (2009). *Recherche d'indicateurs électrodermaux pour l'analyse de la charge mentale en conduite automobile* (thèse de Doctorat, Université Lyon I, France). Récupéré du site thèse en ligne : http://tel.archivesouvertes.fr

Clochard, M. (2011). *L'activité électrodermale, technique pertinente pour l'évaluation des émotions ?* (mémoire de master 2, Université Rennes 1, France). Récupéré du site de l'université Rennes 1 : http://etudes.univ-rennes1.fr/digitalAssets/38/38425_Clochard_activite_electrodermale.pdf

Collet, C., Roure, R., Rada, H., Dittmar, A. et Vernet-Maury, E. (1996). Relationships between performance and skin resistance evolution involving various motor skills. *Physiology and behavior, 59*(4), 953-963.

Critchley, H. D. (2002). Electrodermal Responses: What Happens in the Brain? *The Neuroscientist*, *8*(2), 132.

Darrow, C. W. (1937). Neural Mechanisms Controlling the Palmar Galvanic Skin Reflex and Palmar Sweating: A Consideration of Available Literature. *Archives of Neurology and Psychiatry, 37*(3), 641.

Davies, D. R. et Parasuraman, R. (1982). *The psychology of vigilance*. London, R.-U. : Academic Press.

Dawson, M. E., Schell, A. M. et Filion, D. L. (2007). The electrodermal system. Dans J.T. Cacioppo, L. G. Tassinary et G.G. Berntson (dir.), *Handbook of psychophysiology* (2e éd.) (vol. 2, p. 200-223). New York, NY: Cambridge university press.

Detenber, B. H. et Reeves, B. (1996). A bio-informational theory of emotion: Motion and image size effects on viewers. *Journal of Communication, 46*(3), 66-84.

Detenber, B. H., Simons, R. F. et Bennett, G. G. (1998). Roll 'em!: The effects of picture motion on emotional responses. *Journal of Broadcasting and Electronic Media, 42*(1), 113-127.

Dinges, D. F. et Grace, R. (1998). *Perclos: A valid psychophysiological measure of alertness as assessed by psychomotor vigilance* (Technical Rapport No. FHWA-MCRT- 98-006). U.S. Dept. Transportation, National Highway Traffic Safety Administration.

Dinges, D. F., Mallis, M. M., Maislin, G. et Powell, I. V. (1998). *Evaluation of techniques for ocular measurement as an index of fatigue and the basis for alertness management* (Rapport No. HS-808 762). U.S. Dept. Transportation, National Highway Traffic Safety Administration.

Dixon, N. F. et Spitz, L. (1980). The detection of auditory visual desynchrony. *Perception, 9*(6), 719-721.

Driver, J. (1996). Enhancement of selective listening by illusory mislocation of speech sounds due to lip-reading. *Nature, 381*, 66-68.

Duffy, E. (1972). Activation. Dans N.S. Greenfield et R.A. Sternbach (dir.), *Handbook of psychophysiology* (vol. 5, p. 577-595). New York, NY : Rinehart & Winston.

Durin, V., Gros, L. et Chateau, N. (2006). *Evaluation indirecte de la qualité vocale perçue*. Communication présentée au 8e Congrès Français d'Acoustique (CFA), Tours, France.

Eason, R. G. et Dudley, L. M. (1970). Physiological and behavioral indicants of activation. *Psychophysiology, 7*(2), 223-232.

Edelberg, R. et Wright, D. J. (1964). Two galvanic skin response effector organs and their stimulus specificity. *Psychophysiology, 1*(1), 39-47.

Ekman, P. (1999). Basic emotions. Dans T. Dalgleish et T. Power (dir.), *The Handbook of Cognition and Emotion* (p. 45-60). Sussex, R. -U.: Wiley & Sons.

Ekman, P., Levenson, R. W. et Friesen, W. V. (1983). Autonomic nervous system activity distinguishes among emotions. *Science, 221*, 1208-1210.

Epstein, W. (1975). Recalibration by pairing: a process of perceptual learning. *Perception, 4*(1), 59-72.

Eui Chul, L., Hwan, H. et Kang Ryoung, P. (2010). The comparative measurements of eyestrain caused by 2D and 3D displays. *IEEE transactions Consumer Electronics, 56*(3), 1677-1683.

Eysenck, M. W. (1976). Arousal, learning, and memory. *Psychological bulletin, 83*(3), 389.

Féré, C. (1888). Note sur les modifications de la résistance électrique sous l'influence des excitations sensorielles et des émotions. *Comptes rendus des séances de la société biologique, 5*, 217-219.

Fishel, S. R., Muth, E. R. et Hoover, A. W. (2007). Establishing Appropriate Physiological Baseline Procedures for Real-Time Physiological Measurement. *Journal of Cognitive Engineering and Decision Making, 1*(3), 286-308.

Fort, A. (2002). *Corrélats électrophysiologiques de l'intégration des informations auditives et visuelles dans la perception intermodale chez l'homme.* (thèse de doctorat, Université Lumière Lyon 2, France). Récupéré du site des thèses électroniques de l'université Lumière Lyon 2 : http://theses.univ-lyon2.fr/

Foxe, J. J. et Schroeder, C. E. (2005). The case for feedforward multisensory convergence during early cortical processing. *Neuroreport, 16*(5), 419.

Fowles, D. C., Christie, M. J., Edelberg, R., grings, W. W., Lykken, D. T. et Venables, P. H. (1981). Publication Recommandations for electrodermal Measurements. *Psychophysiology Committee report, 18*(3), 232-239.

Frijda, N. H. (1986). *The emotions*. Cambridge , R. -U. : Cambridge University Press.

Gaillard, A. W. K. (1993). Comparing the concepts of mental load and stress. *Ergonomics, 36*(9), 991-1005.

Geacintov, T. et Peavler, W. S. (1974). Pupillography in industrial fatigue assessment. *Journal of Applied Psychology, 59*(2), 213-216.

Gendolla, G. H. et Krüsken, J. (2001). The joint impact of mood state and task difficulty on cardiovascular and electrodermal reactivity in active coping. *Psychophysiology, 38*(3), 548-556.

Gerin, W., Pieper, C. et Pickering, T. G. (1994). Anticipatory and residual effects of an active coping task on pre-and post-stress baselines. *Journal of psychosomatic research, 38*(2), 139-149.

Gosselin, L. (2003). La fatigue visuelle. *Le Médecin du Québec, 38*(5), 99-102.

Grant, K. W. et Braida, L. D. (1991). Evaluating the articulation index for auditory–visual input. *The Journal of the Acoustical Society of America, 89*, 2952-2960.

Greenwald, M. K., Cook, E. W. et Lang, P. J. (1989). Affective judgment and psychophysiological response: Dimensional covariation in the evaluation of pictorial stimuli. *Journal of psychophysiology, 3*, 51-64.

Groh, J.M., Trause, A.S., Underhill, A.M., Clark, K.R. et Inati, S. (2001). Eye position influences auditory responses in primate inferior colliculus. *Neuron*, *29*(2), 509-518.

Gros, L. (2001). *Evaluation subjective de la qualité vocale fluctuante* (thèse de doctorat non publiée). Université de la méditerranée Aix Marseille II, France.

Gros, L., Chateau, N. et Durin, V. (2006). Speech quality: beyond the MOS score. Dans *Proceedings of the Fifth International Conference on Measurement of Audio and Video Quality in Networks*, (MESAQIN), Prague, République Tchèque.

Gros, L., Chateau, N. et Macé, A. (2005). Assessing speech quality: a new approach. Dans *Proceedings of the Forum Acusticum*, Budapest, Hongrie.

Grossman, P., Stemmler, G. et Meinhardt, E. (1990). Paced respiratory sinus arrhythmia as an index of cardiac parasympathetic tone during varying behavioral tasks. *Psychophysiology, 27*(4), 404-416.

Haber, R. N. et Hershenson, M. (1980). *The psychology of visual perception*. New York, NY: Holt, Rinehart & Winston.

Hancock, P. et Meshkati, N.(1988). *Human mental workload*. Amsterdam, Pays-Bas : Elsevier.

Hands, D. S. (2004). A basic multimedia quality model. *IEEE Transactions on Multimedia, 6*(6), 806-816.

Hastrup, J. L. (1979). Effects of electrodermal lability and introversion on vigilance decrement. *Psychophysiology, 16*(3), 302-310.

Hebb, D. O. (1955). Drives and the CNS (conceptual nervous system). *Psychological Review, 62*(4), 243-254.

Hess, E. H. (1972). Pupillometrics: A method of studying mental, emotional and sensory processes. Dans N. S. Greenfield et R. A. Sternbach (dir.), *Handbook of psychophysiology* (p. 491-531). New York, NY : Rinehart & Winston.

Hess, E. H. et Polt, J. M. (1964). Pupil size in relation to mental activity during simple problem-solving. *Science, 143*, 1190-1192.

Hirsh, I. J. et Sherrick Jr, C. E. (1961). Perceived order in different sense modalities. *Journal of experimental psychology, 62*(5), 423.

Hollier, M. P. et Voelcker, R. M. (1997). Towards a multi-modal perceptual model. *BT Technology Journal, 15*(4), 163-172.

Hollier, M. P. et Rimell, A. N. (1998). An Experimental Investigation into Multi-Modal Synchronisation Sensitivity for Perceptual Model Development. Dans *Proceedings of the 105rd Convention of the Audio Engineering Society* (AES), San Francisco, USA.

Hollier, M. P., Rimell, A. N., Hands, D. S. et Voelcker, R. M. (1999). Multi-modal Perception. *BT Technology Journal, 17*(1), 35-46.

Howard, I. P. (1982). *Human visual orientation*. New York, NY: Wiley.

Howard, I.P. et Templeton, W.B. (1966). *Human spatial orientation.* . New York, NY: J. Wiley.

Hubert, W. et de Jong-Meyer, R. (1991). Autonomic, neuroendocrine, and subjective responses to emotion-inducing film stimuli. *International Journal of Psychophysiology, 11*(2), 131-140.

Ikehara, C. S. et Crosby, M. E. (2005). Assessing cognitive load with physiological sensors. Dans *Proceedings of the 38th Annual International Conference on the System Sciences* (HICSS), 295-303.

ISO/IEC 15938-(2004). *Information Technology - Multimedia Content Description Interface, Part 1–6* (ISO/IEC JTC1/SC29/WG11 N6828). Palma de Majorque, Espagne : MPEG 7.

Jack, C. E. et Thurlow, W. R. (1973). Effects of degree of visual association and angle of displacement on the" ventriloquism" effect. *Perceptual and motor skills, 37*(3), 967-979.

Jackson, C. (1953). Visual factors in auditory localization. *Quarterly Journal of Experimental Psychology, 5*(2), 52-65.

Jae-Hwan, Y., Byoung-Hoon, L. et Deok-Hwan, K. (2012). EOG based eye movement measure of visual fatigue caused by 2D and 3D displays. Dans *Proceedings of the International Conference on Biomedical and Health Informatics* (BHI), 305-308.

James, W. (1884). What is emotion?. *Mind*, *9*, 188-205.

Jennings, J. R., Kamarck, T., Stewart, C., Eddy, M. et Johnson, P. (1992). Alternate cardiovascular baseline assessment techniques: Vanilla or resting baseline. *Psychophysiology, 29*(6), 742-750.

Juris, M. et Velden, M. (1977). The pupillary response to mental overload. *Physiological Psychology*. *5*(4), 421-424.

Just, M. A. et Carpenter, P. A. (1984). Using eye fixations to study reading comprehension. Dans D. E. Kieras et M. A. Just (dir.), *New Methods in Reading Comprehension Research* (p. 151-182). Hillsdale, NJ: Erlbaum.

Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice-Hall.

Kahneman, D. et Peavler, W. S. (1969). Incentive effects and pupillary changes in association learning. *Journal of experimental psychology, 79*, 312-318.

Kahneman, D., Tursky, B., Shapiro, D. et Crider, A. (1969). Pupillary, heart rate, and skin resistance changes during a mental task. *Journal of Experimental Psychology, 79*, 164-167.

Kalsbeek, J. (1971). Sinus arrhythmia and the dual task method in measuring mental load. Dans W. T. Singleton et D. Whitfield (dir.), *Measurement of Man at Work* (p. 101-113). London, R.-U.: Taylor & Francis.

Kamath, M. V. et Fallen, E. (1993). Power spectral analysis of heart rate variability: a noninvasive signature of cardiac autonomic function. *Critical reviews in biomedical engineering, 21*(3), 245.

Kohlrausch, A. et van de Par, S. (1999). Auditory-visual interaction: From fundamental research in cognitive psychology to (possible) applications. Dans B.E. Rogowitz et T. N Pappas. (dir.), *Proceedings of SPIE conference*, *3644*, 34-44.

Kohlrausch, A. et van de Par, S. (2005). Audio-visual interaction in the context of multi-media applications. Dans J. Blauert (dir.), *Communication acoustics* (p. 109-138). Berlin, Allemagne: Springer-Verlag.

Komiyama, S. (1989). Subjective evaluation of angular displacement between picture and sound directions for HDTV sound systems. *Journal of audio engineering society (AES), 37*, 210-214.

King, A. J. (2005). Multisensory integration: strategies for synchronization. *Current biology, 15*(9), 339-341.

Kistler, A., Mariauzouls, C. et von Berlepsch, K. (1998). Fingertip temperature as an indicator for sympathetic responses. *International Journal of Psychophysiology, 29*(1), 35-41.

Klingner, J., Kumar, R. et Hanrahan, P. (2008). Measuring the task-evoked pupillary response with a remote eye tracker. Dans *Proceedings of the symposium on Eye tracking research and applications* (ETRA), 69-72.

Knoche, H., De Meer, H. G. et Kirsh, D. (1999). Utility curves: Mean opinion scores considered biased. Dans *Proceedings of the 7th International Workshop on the Quality of Service* (IWQoS), 12-14.

Koch, C. (2004). *The Quest for Consciousness: A Neurobiological Approach* . Englewood, Colorado : Robert & Company Publishers.

Kramer, A. F. (1991). Physiological metrics of mental workload: A review of recent progress. Dans L. Damos (dir.), *Multiple-task performance* (p. 279-328). London, R.-U. : Taylor & Francis.

Laboratoire D'anthropologie Appliquée (1996). *Mise en place d'une méthode d'étude de la fatigue des pilotes dans le transport aérien, phase 1* (Rapport AA 358/96). Paris, France. Récupéré de http://www.developpement-durable.gouv.fr/IMG/pdf/fatigue1.pdf

Lacey, J. I. (1967). Somatic response patterning and stress: Some revisions of activation theory. Dans M. Appley et R. Trumbull (dir.), *Psychological stress: Issues in research* (p. 14-42). New York, NY: Appleton century crofts.

Lacey, J. I. et Lacey, B. C. (1958). Verification and extension of the principle of autonomic response-stereotypy. *The American Journal of Psychology, 71*(1), 50-73.

Lacey, J. I. et Lacey, B. C. (1970). Some autonomic-central nervous system interrelationships. Dans P. Black (dir.), *Physiological correlates of emotion* (p. 205-227). New York, NY: Academic Press.

Lacombe, M. (2009). *Lacombe : précis d'anatomie et de physiologie humaines* (vol. 2, 30e éd.). Rueil-Malmaison, France: Lamarre.

Lambooij, M., Fortuin, M., Heynderickx, I. et IJsselsteijn, W. (2009). Visual discomfort and visual fatigue of stereoscopic displays: a review. *Journal of Imaging Science, 53*(3), 1-14.

Lang, A. (1990). Involuntary attention and physiological arousal evoked by structural features and emotional content in TV commercials. *Communication Research, 17*(3), 275-299.

Lang, A. (1991). Emotion, formal features, and memory for televised political advertisements. *Television and political advertising, 1*, 221-243.

Lang, A. (1995). Defining audio/video redundancy from a limited-capacity information processing perspective. *Communication Research, 22*(1), 86-115.

Lang, A. et Basil, M. (1998). Attention, resource allocation, and communication research: What do secondary task reaction times measure, anyway? In M. Roloff (dir.), *Mass Communication yearbook* (vol. 21, p. 443-473). Beverly Hills, CA: Sage.

Lang, A., Bolls, P., Potter, R. F. et Kawahara, K. (1999). The effects of production pacing and arousing content on the information processing of television messages. *Journal of Broadcasting and Electronic Media, 43*(4), 451-475.

Lang, A., Newhagen, J. et Reeves, B. (1996). Negative video as structure: Emotion, attention, capacity, and memory. *Journal of Broadcasting and Electronic Media, 40*(4), 460-477.

Lang, A., Zhou, S., Schwartz, N., Bolls, P. D. et Potter, R. F. (2000). The effects of edits on arousal, attention, and memory for television messages: When an edit is an edit can an edit be too much? *Journal of Broadcasting and Electronic Media, 44*(1), 94-109.

Lang, L. et Qi, H. (2008). The study of driver fatigue monitor algorithm combined PERCLOS and AECS. Dans *Proceedings of the International Conference on Computer Science and Software Engineering*, *1*, 349-352.

Lang, P. J. (1980). Behavioral treatment and bio-behavioral assessment: Computer applications. Dans J. B. Sidowski, J. H. Johnson et G. A. Williams (dir.), *Technology in mental health care delivery systems* (p. 119-137). Norwood, NJ: Ablex.

Lang, P. J., Greenwald, M. K., Bradley, M. M. et Hamm, A. O. (1993). Looking at pictures: Affective, facial, visceral, and behavioral reactions. *Psychophysiology, 30*(3), 261-273.

Lantelme, P., Custaud, M. A., Vincent, M. et Milon, H. (2002). Implications cliniques de la variabilité tensionnelle. *Archives des maladies du coeur et des vaisseaux, 95*(9), 787-792.

Lazarus, R. S., Speisman, J. C., Mordkoff, A. M. et Davison, L. A. (1962). A laboratory study of psychological stress produced by a motion picture film. *Psychological Monographs: General and Applied*, *76*(34), 1-35.

Libby Jr, W. L., Lacey, B. C. et Lacey, J. I. (1973). Pupillary and cardiac activity during visual attention. *Psychophysiology, 10*(3), 270-294.

Lin, T., Imamiya, A., Hu, W. et Omata, M. (2007). Combined user physical, physiological and subjective measures for assessing user cost. *Universal access in ambient intelligence environments,* 304-316.

Lin, T., Omata, M., Hu, W. et Imamiya, A. (2005). Do physiological data relate to traditional usability indexes? Dans *Proceedings of the 17th Australia conference on Computer-Human Interaction*: *Citizens Online: Considerations for Today and the Future*, 1-10.

Lindsley D.B. (1951). Emotions. Dans S. S. Stevens (dir.), *Handbook of Experimental Psychology* (p. 473-516). New York, NY : Wiley.

Lombard, M. (1995). Direct Responses to People on the Screen Television and Personal Space. *Communication Research, 22*(3), 288-324.

Lowenstein, O. et Loewenfeld, I. E. (1964). The slepp-waking cycle and pupillary activity. *Annals of the New York Academy of Sciences, 117*(1), 142-156.

Lussier, F. et Flessas, J. (2001). *Neuropsychologie de l'enfant: troubles développementaux et de l'apprentissage*. Paris, France : Dunod.

Malliani, A. (1999). The Pattern of Sympathovagal Balance Explored in the Frequency Domain. *News in physiological sciences*, *14*, 111-117.

Malliani, A., Lombardi, F., Pagani, M. et Cerutti, S. (1990). Clinical exploration of the autonomic nervous system by means of electrocardiography. *Annals of the New York Academy of Sciences, 601*(1), 234-246.

Malliani, A., Pagani, M., Lombardi, F. et Cerutti, S. (1991). Cardiovascular neural regulation explored in the frequency domain. *Circulation, 84*(2), 482-492.

Mandryk, R. L., Inkpen, K. M. et Calvert, T. W. (2006). Using psychophysiological techniques to measure user experience with entertainment technologies. *Behaviour et information technology, 25*(2), 141-158.

Mascetti, G. G. et Strozzi, L. (1988). Visual cells in the inferior colliculus of the cat. *Brain research, 442*(2), 387-390.

Massaro, D. W., Cohen, M. M. et Smeele, P. M. T. (1996). Perception of asynchronous and conflicting visual and auditory speech. *The Journal of the Acoustical Society of America, 100*(3), 1777-1786.

May, J. G., Kennedy, R. S., Williams, M. C., Dunlap, W. P. et Brannan, J. R. (1990). Eye movement indices of mental workload. *Acta psychologica, 75*(1), 75-89.

McGurk, H. et MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264*, 746-748.

Meehan, M., Insko, B., Whitton, M. et Brooks Jr, F. P. (2002). *Physiological measures of presence in stressful virtual environments. ACM Transactions on Graphics, 21*(3), 645-652.

Menche, N. et Schäffler, A. (2004). *Anatomie physiologie biologie: abrégé d'enseignement pour les professions de santé* (2e éd.). Paris, France : Maloine.

Mills, A. W. (1958). On the minimum audible angle. *The Journal of the Acoustical Society of America, 30*(4), 237-246.

Montano, N., Ruscone, T. G., Porta, A., Lombardi, F., Pagani, M. et Malliani, A. (1994). Power spectrum analysis of heart rate variability to assess the changes in sympathovagal balance during graded orthostatic tilt. *Circulation, 90*(4), 1826-1831.

Mordkoff, A. M. (1964). The relationship between psychological and physiological response to stress. *Psychosomatic medicine, 26*(2), 135-150.

Morris, T. et Miller, J. (1996). Electrooculographic and performance indices of fatigue during simulated flight. *Biological Psychology, 42*(3), 343-360.

Mourant, R. R., Lakshmanan, R. et Chantadisai, R. (1981). Visual fatigue and cathode ray tube display terminals. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 23*(5), 529-540.

Mullin, J., Smallwood, L., Watson, A. et Wilson, G. M. (2001). New techniques for assessing audio and video quality in real-time interactive communications. Dans Vanderdonckt, A., A. blandford et A., Derycke (dir.), *Proceedings of IHM HCI*, Lille, France.

Nakayama, M., Takahashi, K. et Shimizu, Y. (2002). The act of task difficulty and eye-movement frequency for the'Oculo-motor indices. Dans *Proceedings of the Symposium on Eye tracking research and applications* (ETRA), 37-42.

Norman, D. A. et Bobrow, D. G. (1975). On data-limited and resource-limited processes. *Cognitive psychology, 7*(1), 44-64.

Obrist, P. A. (1981). *Cardiovascular psychology: A perspective.* New York, NY: Plenum.

Olive, T., Piolat, A. et Roussey, J.-Y. (1997). Effort cognitif et mobilisation des processus en production de texte: effet de l'habileté rédactionnelle et du niveau de connaissances. *Attention et contrôle cognitif: Mécanismes, développement des habiletés, pathologies*, 71-85.

Orchard, L. N. et Stern, J. A. (1991). Blinks as an index of cognitive activity during reading. *Integrative physiological and behavioral science, 26*(2), 108-116.

Pagani M, Furlan R, Pizzinelli P, Crivellaro, W, Cerutti S, Malliani A. (1989). Spectral analysis of R-R and arterial pressure variabilities to assess sympathovagal interaction during mental stress in humans. *Journal of Hypertension*, 7(6).

Pagani, M., Lombardi, F., Guzzetti, S., Rimoldi, O., Furlan, R., Pizzinelli, P., …Piccaluga, E. (1986). Power spectral analysis of heart rate and arterial pressure variabilities as a marker of sympatho-vagal interaction in man and conscious dog. *Circulation research, 59*(2), 178-193.

Palhais, J., Cruz, R. et Nunes, M. (2012). Quality of Experience Assessment in Internet TV. Dans K. Pentikousis, R. Aguiar, S. Sargento et R. Agüero (dir.), *Mobile Networks and Management* (vol. 97, p. 261-274). Berlin Heidelberg : Springer.

Paloff, A. et Usunoff, K. (1992). Projections to the inferior colliculus from the dorsal column nuclei. An experimental electron microscopic study in the cat. *Journal für Hirnforschung, 33*(6), 597.

Partala, T., Jokiniemi, M. et Surakka, V. (2000). Pupillary responses to emotionally provocative stimuli. Dans *Proceedings of the Symposium on Eye tracking research and applications* (ETRA), 123-129.

Pastrana-Vidal, R. R. (2005). *Vers une métrique perceptuelle de qualité audiovisuelle dans un contexte à service non garanti.* (thèse de doctorat non publiée). Université de Bourgogne, France.

Paulhus, D. L. (2002). Socially desirable responding: The evolution of a construct. Dans Braun, H. I., Jackson, D. N. (dir.), *The role of constructs in psychological and educational measurement* (p.37-48). Mahwah NJ: Erlbaum.

Porges, S. W. (1992). Vagal tone: a physiologic marker of stress vulnerability. *Pediatrics, 90*(3), 498-504.

Porges, S. W. (1995). Orienting in a defensive world: Mammalian modifications of our evolutionary heritage : a polyvagal theory. *Psychophysiology, 32*(4), 301-318.

Porges, S. W. et Byrne, E. A. (1992). Research methods for measurement of heart rate and respiration. *Biological Psychology, 34*(2), 93-130.

Porter, G., Troscianko, T. et Gilchrist, I. D. (2007). Effort during visual search and counting: Insights from pupillometry. *The Quarterly Journal of Experimental Psychology, 60*(2), 211-229.

Posner, M. I., Rueda, M. R. et Kanske, P. (2007). Probing the mechanisms of attention. Dans J.T. Cacioppo, J.G. Tassinary et G.G. Berntson (dir.), *Handbook of Psychophysiology* (p. 410-432). Cambridge, R. -U.: Cambridge University Press.

Pribram, K. H. et McGuinness, D. (1975). Arousal, activation, and effort in the control of attention. *Psychological Review, 82*(2), 116-149.

Radeau, M. (1994). Auditory-visual spatial interaction and modularity. *Cahiers de Psychologie Cognitive, 13*, 3-51.

Radeau, M. et Bertelson, P. (1977). Adaptation to auditory-visual discordance and ventriloquism in semi-realistic situations. *Attention, Perception et Psychophysics, 22*(2), 137-146.

Radeau, M. et Bertelson, P. (1978). Cognitive factors and adaptation to auditory-visual discordance. *Attention, Perception et Psychophysics, 23*(4), 341-343.

Ragot, R., Cavé, C. et Fano, M. (1988). Reciprocal effects of visual and auditory stimuli in a spatial compatibility situation. *Bulletin of the Psychonomic Society, 26*, 350-352.

Rani, P., Sims, J., Brackin, R. et Sarkar, N. (2002). Online stress detection using psychophysiological signals for implicit human-robot cooperation. *Robotica, 20*(06), 673-685.

Ravaja, N. (2004). Contributions of psychophysiology to media research: Review and recommendations. *Media Psychology, 6*(2), 193-235.

Ravaja, N., Saari, T., Laarni, J., Kallinen, K., Salminen, M., Holopainen, J. et Järvinen, A. (2005). The psychophysiology of video gaming: Phasic emotional responses to game events. Dans *Proceedings of the DiGRA conference Changing views: worlds in play*. Vancouver, Canada.

Rayner, K. (1998). Eye movements in reading and information processing : 20 years of research. *Psychological bulletin, 124*(3), 372.

Reeves, B., Lang, A., Kim, E. Y. et Tatar, D. (1999). The effects of screen size and message content on attention and arousal. *Media Psychology, 1*(1), 49-67.

Reeves, B., Thorson, E., Rothschild, M. L., McDonald, D., Hirsch, J. et Goldstein, R. (1985). Attention to television: Intrastimulus effects of movement and scene changes on alpha variation over time. *International Journal of Neuroscience*, *27*(3-4), 241-255.

Reiter, U. et Weitzel, M. (2007). Influence of Interaction on Perceived Quality in Audiovisual Applications: Evaluation of Cross-Modal Influence. Dans *Proceedings of the 13th International Conference on Auditory Displays* (ICAD), Montreal, Canada.

Reiter, U., Weitzel, M. et Cao, S. (2007). Influence of Interaction on Perceived Quality in Audio Visual Applications: Subjective Assessment with n-Back Working Memory Task. Dans *Proceedings of the 30th AES International Conference on intelligent Audio Environments*, Saariselk, Finland.

Remington, R. (1980). Attention and saccadic eye movements. *Journal of Experimental Psychology: Human Perception and Performance, 6,* 726-744.

Rihs, S. (1995). The influence of audio on perceived picture quality and subjective audio-video delay tolerance. Dans R. Hamber et H. Ridder (dir.), *Proceeding of the MOSAIC workshop: Advanced Methods for the Evaluation of Television Picture Quality* (p. 133-137). Eindhoven, Netherlands: IPO.

Rimell, A. N. et Owen, A. (2000). The effect of focused attention on audio-visual quality perception with applications in multi-modal codec design. Dans *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing* (ICASSP), *4*, 2377-6293.

Roscoe, A. H. (1992). Assessing pilot workload. Why measure heart rate, HRV and respiration? *Biological Psychology, 34*(2), 259-287.

Rowe, D. W., Sibert, J. et Irwin, D. (1998). Heart rate variability: Indicator of user state as an aid to human-computer interaction. Dans *Proceedings of the SIGCHI conference on Human factors in computing systems*, 480-487.

Salahuddin, L., Cho, J., Jeong, M. G. et Kim, D. (2007). Ultra short term analysis of heart rate variability for monitoring mental stress in mobile settings. Dans *Proceedings of the 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (EMBS), 4656-4659.

Sanders, A. (1990). Issues and trends in the debate on discrete vs. continuous processing of information. *Acta psychologica, 74*(2), 123-167.

Schmidt, D., Abel, L., DellOsso, L. et Daroff, R. (1979). Saccadic velocity characteristics-Intrinsic variability and fatigue. *Aviation, Space, and Environmental Medicine, 50*(4), 393-395.

Sekuler, R. et Blake, R. (1990). *Perception*. (2e éd.). New-York, NY : McGraw-Hill.

Sekuler, R., Sekuler, A. B. et Lau, R. (1997). Sound alters visual motion perception. *Nature, 385*(6614), 308.

Shi, Y., Ruiz, N., Taib, R., Choi, E. et Chen, F. (2007). Galvanic skin response (GSR) as an index of cognitive load. Dans *Proceedings of the conference of Computer Humain Interaction* (CHI), 2651-2656.

Siddle, D. A. T. (1991). Orienting, habituation, and resource allocation: An associative analysis. *Psychophysiology, 28*(3), 245-259.

Simons, R. F., Detenber, B. H., Roedema, T. M. et Reiss, J. E. (1999). Emotion processing in three systems: The medium and the message. *Psychophysiology, 36*(5), 619-627.

Sommer, D. et Golz, M. (2010). Evaluation of PERCLOS based current fatigue monitoring technologies. Dans *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (EMBC), 4456 – 4459.

Sostek, A. J. (1978). Effects of electrodermal lability and payoff instructions on vigilance performance. *Psychophysiology, 15*(6), 561-568.

Souza Neto, E., Neidecker, J. et Lehot, J. (2003). *Comprendre la variabilité de la pression artérielle et de la fréquence cardiaque. Annales françaises d'anesthésie et de réanimation*, *22*, 425-452.

Spence, C. (2007). Audiovisual multisensory integration. *Acoustical science and technology, 28*(2), 61-70.

Stein, B. E. et Meredith, M. A. (1993). *The merging of the senses*. Massachussets, MA : The MIT Press.

Stein, B. E., Wallace, M. T. et Meredith, M. A. (1995). Neural mechanisms mediating attention and orientation to multisensory cues. Dans M. Gazzaniga (dir.), *The cognitive neurosciences* (p. 683-702). Cambridge, MA : The MIT Press.

Stern, J.A. (1980). *Aspects of visual search activity related to attentional processes and skill development* (Rapport n° F49620–79-C0089). Washington, DC: Air Force Office of Scientific Research.

Stern, J. A., Boyer, D. et Schroeder, D. (1994). Blink rate: a possible measure of fatigue. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 36*(2), 285-297.

Stern, J. A., Boyer, D., Schroeder, D., Touchstone, M. et Stoliarov, N. (1994). *Blinks, Saccades, and Fixation Pauses During Vigilance Task Performance: 1* (rapport n° DOT/FAA/AM94/26). Washington, DC: Office of Aviation Medicine.

Stern, J. A. et Dunham, D. N. (1990). The ocular system. DansJ. T. Cacioppo et L. G. Tassinary (dir.), *Principles of psychophysiology* (p. 193-215). Cambridge : Cambridge University Press.

Stern, R. M., Ray, W. J. et Quigley, K. S. (2001). *Psychophysiological recording* (2e éd.). New York, NY : Oxford University Press.

Sumby, W. H. et Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *The Journal of the Acoustical Society of America, 26*(2), 212-215.

Summerfield, Q. (1979). Use of visual information for phonetic perception. *Phonetica, 36*(4-5), 314-331.

Task force of the European Society of Cardiology and the North America Society of pacing and electrophysiology (1996). Heart rate variability: standards of measurement, physiological interpretation, and clinical use. *European Heart Journal, 17*, 354–381.

Tattersall, A. J. et Hockey, G. R. J. (1995). Level of operator control and changes in heart rate variability during simulated flight maintenance. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 37*(4), 682-698.

Tecce, J. (1992). Psychology, physiology and experimental. Dans *McGraw-Hill Yearbook of Science and Technology* (p. 375-377). New York, NY: McGraw-Hill.

Thong, T., Li, K., McNames, J., Aboy, M. et Goldstein, B. (2003). Accuracy of ultra-short heart rate variability measures. Dans *Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (EMBS), *3*, 2424-2427.

Thorson, E. et Lang, A. (1992). The effects of television videographics and lecture familiarity on adult cardiac orienting responses and memory. *Communication Research, 19*(3), 346-369.

Thurlow, W. R. et Jack, C. E. (1973). Certain determinants of the ventriloquism effect. *Perceptual and motor skills, 36*(3c), 1171-1181.

Tuch, A. N., Kreibig, S. D., Roth, S. P., Bargas-Avila, J. A., Opwis, K. et Wilhelm, F. H. (2011). The Role of Visual Complexity in Affective Reactions to Webpages: Subjective, Eye Movement, and Cardiovascular Responses. *IEEE Transactions on Affective Computing, 2*(4), 230-236.

Turpin, G., Schaefer, F. et Boucsein, W. (1999). Effects of stimulus intensity, risetime, and duration on autonomic and behavioral responding: Implications for the differentiation of orienting, startle, and defense responses. *Psychophysiology*, *36*(4), 453-463.

UIT- R. (1995). *Evaluation of the subjective effects of timing errors between sound and vision signals in television.* Recommendation SG11 11A/55. Récupéré du site de l'Union Internationale des Télécommunications : http://www.itu.int/en/publications/ITU-R

UIT-R (1997). *Méthodes d'évaluation subjective des dégradations faibles dans les systèmes audio y compris les systèmes sonores multivoies*. Recommandation BS.1116-1. Récupéré du site de l'Union Internationale des Télécommunications : http://www.itu.int/en/publications/ITU-R

UIT-R. (1997). *Methods for the subjective assessment of audio systems with accompanying picture*. Recommandation BS.1286. Récupéré du site de l'Union Internationale des Télécommunications http://www.itu.int/en/publications/ITU-R

UIT-R. (1998). *Synchronisation relative du son et de l'image en radiodiffusion*. Recommandation BT.1359-1. Récupéré du site de l'Union Internationale des Télécommunications : http://www.itu.int/en/publications/ITU-R

UIT-R. (2003). *Méthode d'évaluation subjective du niveau de qualité intermédiaire des systèmes de codage.* Recommandation BS.1534-1. Récupéré du site de l'Union Internationale des Télécommunications : http://www.itu.int/en/publications/ITU-R

UIT-R. (2003). *Méthodes générales d'évaluation subjective de la qualité du son.* Recommandation BS.1284-1. Récupéré du site de l'Union Internationale des Télécommunications : http://www.itu.int/en/publications/ITU-R

UIT-R. (2007). *Méthode d'évaluation subjective de la qualité vidéo dans les applications multimédias.* Recommandation BT.1788. Récupéré du site de l'Union Internationale des Télécommunications : http://www.itu.int/en/publications/ITU-R

UIT-R (2007). *Spécifications et méthodes de réglage de la brillance et du contraste des écrans*. Recommandation BT.814-2. Récupéré du site de l'Union Internationale des Télécommunications : http://www.itu.int/en/publications/ITU-R

UIT-R. (2012). *Methodology for the subjective assessment of the quality of television pictures.* Recommendation BT.500-13. Récupéré du site de l'Union Internationale des Télécommunications : http://www.itu.int/en/publications/ITU-R

UIT-R. (2012). *Système de son stéréophonique multicanal avec ou sans image associée*. Recommandation BS.775-3. Récupéré du site de l'Union Internationale des Télécommunications : http://www.itu.int/en/publications/ITU-R

UIT-T (1990). *Tolérances pour la différence entre les temps de transmission des composante son et image d'un signal de télévision*. Recommandation J.100. Récupéré du site de l'Union Internationale des Télécommunications : http://www.itu.int/pub/T-REC

UIT -T. (1996). *Methods for subjective determination of transmission quality.* Recommandation P.800. Récupéré du site de l'Union Internationale des Télécommunications : http://www.itu.int/pub/T-REC

UIT-T (1997). *Relation between audio, video and audiovisual quality* (rapport SG12 COM 12-19-E). Pays-Bas : KPN

UIT-T. (1998). *Méthodes d'évaluation subjective de la qualité audiovisuelle pour applications multimédias*. Recommandation P.911. Récupéré du site de l'Union Internationale des Télécommunications : http://www.itu.int/pub/T-REC

ITU-T (1998). *Study Of The Influence Of Experimental Context On The Relationship Between Audio, Video And Audiovisual Subjective Qualities* (rapport SG 12 COM 12-61 E).

UIT-T. (1999). *Méthodes d'évaluation subjective de la qualité audiovisuelle pour applications multimédias: Corrigendum 1*. Recommandation P.911. Récupéré du site de l'Union Internationale des Télécommunications : http://www.itu.int/pub/T-REC

UIT-T. (1999). *Subjective video quality assessment methods for multimedia applications*. Recommendation P.910. Récupéré du site de l'Union Internationale des Télécommunications : http://www.itu.int/pub/T-REC

UIT-T. (2000). *Méthodes d'essai interactives pour communications audiovisuelles*. Recommandation P.920. Récupéré du site de l'Union Internationale des Télécommunications : http://www.itu.int/pub/T-REC

UIT-T (2007). *Subjective evaluation of conversational quality*. Recommandation P.805. Récupéré du site de l'Union Internationale des Télécommunications : http://www.itu.int/pub/T-REC

UIT-T. (2008). *Définition de termes relatifs à la qualité de service.* Recommandation E.800. Récupéré du site de l'Union Internationale des Télécommunications : http://www.itu.int/pub/T-REC

UIT -T. (2008). *Vocabulary for performance and quality of service, Amendment 2: New definitions for inclusion in Recommendation ITU-T P.10/G.100.* Recommandation P.10/G.100. Récupéré du site de l'Union Internationale des Télécommunications : http://www.itu.int/pub/T-REC

Ukai, K. et Howarth, P. A. (2008). Visual fatigue caused by viewing stereoscopic motion images: Background, theories, and observations. *Displays 29 Health and Safety Aspects of Visual Displays*, *9*(2), 106–116.

VaezMousavi, S. M., Barry, R. J., Rushby, J. A. et Clarke, A. R. (2007). Arousal and activation effects on physiological and behavioral responding during a continuous performance task. *Acta neurobiologiae experimentalis, 67*(4), 461-470.

Vahedian, A., Frater, M. R. et Arnold, J. F. (1999). Impact of audio on subjective assessment of video quality. Dans *Proceedings of the 6th International Conference on Image Processing (ICIP)*, *2*, 367-370.

Van de Par, S., Kohlrausch, A. et Juola, J. F. (2002). Some methodological aspects for measuring asynchrony detection in audio–visual stimuli. Dans *Proceedings of Forum Acusticum, 32*.

Vatakis, A. et Spence, C. (2006). Audiovisual synchrony perception for speech and music assessed using a temporal order judgment task. *Neuroscience letters, 393*(1), 40-44.

Veltman, J. A. et Gaillard, A.W. K. (1998). Physiological workload reactions to increasing levels of task difficulty. *Ergonomics, 41*(5), 656-669.

Venables, P. H. et Christie, M. J. (1980). Electrodermal activity. Dans I. Martin et P. H. Venables (dir.), *Techniques in psychophysiology* (p. 3-67). New York, NY : Wiley & Sons.

Vicente, K. J., Thornton, D. C. et Moray, N. (1987). Spectral analysis of sinus arrhythmia: A measure of mental effort. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 29*(2), 171-182.

Wagner, H. (1993). Sound-localization deficits induced by lesions in the barn owl's auditory space map. *The Journal of Neuroscience, 13*(1), 371-386.

Ward, R. D. et Marsden, P. H. (2003). Physiological responses to different WEB page designs. *International Journal of Human-Computer Studies, 59*(1), 199-212.

Ward, R., Marsden, P., Cahill, B. et Johnson, C. (2002). Physiological responses to well-designed and poorly designed interfaces. *International Journal of Human-Computer Studies*, *59*(1/2), 199-212.

Wastell, D. G. et Newman, M. (1996). Stress, control and computer system design: a psychophysiological field study. *Behaviour and information technology, 15*(3), 183-192.

Welch, R. B., DutionHurt, L. D. et Warren, D. H. (1986). Contributions of audition and vision to temporal rate perception. *Attention, Perception et Psychophysics, 39*(4), 294-300.

Welch, R. B. et Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological bulletin, 88*(3), 638-667.

Welch, R. B. et Warren, D. H. (1989). Intersensory interaction. Dans K. R. Boff, L. Kaufman et J. P. Thomas (dir.), *Handbook of perception and human performance*. Dordrecht, Pays-Bas: Kluwer academic.

Widmaier, E. P., Raff, H. et Strang, K. T. (2012). *Vander physiologie humaine : les mécanismes du fonctionnement de l'organisme* (5e éd.). Paris, France.

Wierwille, W. W., Wreggit, S. S. et Knipling, R. R. (1994). Development of improved algorithms for on-line detection of driver drowsiness. Dans *Proceedings of the International Congress on Transportation Electronics*, 331-340.

Wilcott, R. (1967). Arousal sweating and electrodermal phenomena. *Psychological bulletin, 67*(1), 58-72.

Wilson, G. F. (2002). An analysis of mental workload in pilots during flight using multiple psychophysiological measures. *The International Journal of Aviation Psychology, 12*(1), 3-18.

Wilson, G. F. et Eggemeier, F. T. (1991). Psychophysiological assessment of workload in multi-task environments. *Multiple-task performance*, 329-360.

Wilson, G. F. et O'Donnell, R.D. (1988). Measurement of operator workload with the neurophysiological workload test battery Dans P.A. Hancock et N. Meshkati (dir.), *Human mental workload* (p. 63-100). Amsterdam, Pays-Bas.

Wilson, G.M. et Sasse, M.A. (2000a). Do Users Always Know What's Good For Them? Utilizing Physiological Responses to Assess Media Quality. Dans *Proceedings of Human Computer Interaction (HCI) : People and Computers XIV - Usability or Else*! 327-339.

Wilson, G.M. et Sasse, M.A (2000b). Investigating the Impact of Audio Degradations on Users: Subjective vs. Objective Assessment Methods. Dans *Proceedings of OZCHI : Interfacing Reality in the New Millennium*, 135-142.

Winton, W. M., Putnam, L. E. et Krauss, R. M. (1984). Facial and autonomic manifestations of the dimensional structure of emotion. *Journal of Experimental Social Psychology, 20*(3), 195-216.

Woodmansee, J. (1967). The pupil reaction as an index of positive and negative affect. Dans *Proceedings of the Annual Convention of the American Psychological Association*, Washington, DC.

Woodworth, R. S. et Schlosberg, H. (1954). *Experimental psychology*. New York, NY: Holt, Rinehart & Winston.

Wu, S. et Lin, T. (2011). Exploring the use of physiology in adaptive game design. Dans *Proceedings of the International Conference on the Consumer Electronics, Communications and Networks* (CECNet), 1280-1283.

Wundt W (1896): *Gundriss der Psychologie* [Outlines of Psychology]. Leipzig, Germany: Entgelman.

Yano, S., Ide, S., Mitsuhashi, T. et Thwaites, H. (2002). A study of visual fatigue and visual comfort for 3D HDTV/HDTV images. *Displays, 23*(4), 191-201.

Yerkes, R. M. et Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of comparative neurology and psychology, 18*(5), 459-482.

Yoon, K., Bolls, P. et Lang, A. (1998). The effects of arousal on liking and believability of commercials. *Journal of Marketing Communications, 4*(2), 101-114.

You, J., Reiter, U., Hannuksela, M., Gabbouj, M. et Perkis, A. (2010). Perceptual-based quality assessment for audio–visual services: A survey. *Signal Processing: Image Communication, 25*, 482-501.

REFERENCES COMPLEMENTAIRES

Averty, P. (1998). *Les effets de la charge de trafic sur le niveau d'activation psychophysiologique du contrôleur aérien* (thèse de doctorat non publiée). Université Lumière Lyon 2, France.

Barrouillet, P. (1996). Ressources, capacités cognitives et mémoire de travail: postulats, métaphores et modèles: La charge mentale. *Psychologie française*, *41*(4), 319-338.

Barry, R. J. (2004). Stimulus significance effects in habituation of the phasic and tonic orienting reflex. *Integrative physiological and behavioral science, 39*(3), 166-179.

Caldwell, J., Wilson, G., Cetinguc, M., Gaillard, A., Gundel, A., Lagarde, D., …Wright, A. (1994). *Psychophysiological assessment methods*. (report AGARD-AR vol.324). Neuilly-sur-Seine, France : NATO Advisory Group for Aerospace Research and Development.

Collet, C., Averty, P. et Dittmar, A. (2009). Autonomic nervous system and subjective ratings of strain in air-traffic control. *Applied Ergonomics*, *40*(1), 23-32. doi: 10.1016/j.apergo.2008.01.019

Collet, C., Petit, C., Champely, S. et Dittmar, A. (2003). Assessing workload through physiological measurements in bus drivers using an automated system during docking. *Journal of the Human Factors and Ergonomics Society, 45*(4), 539-548.

Desnoyers, L. (1987). *Travail visuel, fatigue visuelle*. France: C.N.A.M.

Farha, J. G. et Sher, K. J. (1989). The effects of consent procedures on the psychophysiological assessment of anxiety: A methodological inquiry. *Psychophysiology, 26*(2), 185-191.

Gaillard, A. W. K. et Kramer, A. F. (2000). Theoretical and methodological issues in psychophysiological research. DansR. W. Backs et W. Boucsein (dir.), *Engineering psychophysiology: Issues and applications* (p. 31-58). Mahwah, New Jersey, USA: Lawrence Erlbaum Associates.

Goliot-Lété, A. et Vanoye, F. (1993). *Précis d'analyse filmique*. Paris, France : Nathan Université.

Hancock, P. (1986). Stress and adaptability. Dans G. R. J. Hockey, A. W. K. Gaillard et M. G. H. Coles (dir.), *Energetics and human information processing* (p. 243-251). Dordrecht, Pays-Bas : Springer.

Hockey, G. R. J., Coles, M. G. et Gaillard, A. W. K (1986). Energetical issues in research on human information processing. Dans G. R. J. Hockey, A. W. K. Gaillard et M. G.

Coles (dir.), *Energetics and human information processing* (p. 3-21). Dordrecht, Pays-Bas: Springer.

Jones, B.L. et McManus, P.R. (1986). Graphic scaling of qualitative terms. *SMPTE Journal*, 1166-1171.

Mitchell, D. K. (2000). Mental workload and ARL workload modeling tools (report n° ARL-TN-161). Aberdeen Proving Ground, MD : U.S. Army Research Laboratory.

Nikula, R. (1991). Psychological Correlates of Nonspecific Skin Conductance Responses. *Psychophysiology*, *28*(1), 86-90. doi: 10.1111/j.1469-8986.1991.tb03392.x

Ponder, E. et Kennedy, W. (1927). On the act of blinking. *Experimental Physiology, 18*(2), 89-110.

Posner, M. L et Rothbaft, M. K. (2004). Hebb's Neural networks support the integration of psychological science. *Canadian Psychologist*, *45*, 265-278.

Sweeney, M., Maguire, M. et Shackel, B. (1993). Evaluating user-computer interaction: a framework. *International Journal of Man-Machine Studies, 38*(4), 689-711.

Turner, J. R. (1994). *Cardiovascular Reactivity and Stress: Patterns of Physiological Response*. New York, NY: Springer.

VaezMousavi, S. M., Hashemi-Masoumi, E. et Jalali, S. (2008). Arousal and Activation in a Sport Shooting Task. *World Applied Sciences Journal*, *4*(6), 824-829.

Virtanen, M.T., Gleiss, N. et Goldstein, M. (1995). On the use of evaluative category scales in telecommunications. Dans *Proceedings of Human Factors in Telecommunications*, *95*, 253-260.

Zettl, H. (1991). *Sight, sound, motion: Applied media aesthetics*. Belmont, CA: Wadsworth.

ANNEXE 2-A

Système perceptif audiovisuel

Les descriptions suivantes ne se veulent pas exhaustives mais proposent une approche succincte des bases fonctionnelles et anatomophysiologiques des systèmes auditifs et visuels humain.

A. SYSTEME AUDITIF

L'information auditive se définit comme l'addition des ondes sonores issues, de manière simultanée, des différentes sources provenant du monde extérieur. La scène auditive résultante va être présentée à l'oreille puis analysée pour permettre son interprétation. Les paragraphes suivants détailleront de manière succincte les différentes étapes du traitement de l'information auditive du point de vue anatomique, physiologique et fonctionnel.

A.1 SIGNAL PHYSIQUE

L'énergie sonore peut être définie comme des variations de pression propagées, à une vitesse finie et assujettie au milieu ambiant (380m/s dans l'air contre 1480 m/s dans l'eau), grâce aux vibrations de molécules de proche en proche. En absence de molécule (milieu matériel), comme dans le vide, le son ne peut se propager. Tout objet ou phénomène capable d'entraîner une perturbation moléculaire peut constituer une source sonore. Ces vibrations se

diffusent de manière ondulatoire se caractérisant par la fréquence (la vitesse de vibration, en Hertz), l'amplitude (force de la pression produite) et la phase (son décalage dans le temps). Plus l'amplitude est grande, plus le son sera fort, plus la fréquence est élevée, plus le son sera aigu. Le système auditif humain peut percevoir un spectre auditif compris entre 20 et 20 000 Hz avec une acuité optimale entre 1000 et 4000 Hz. En dehors de ces bornes, le son est inaudible (infra et ultrasons) ou dangereux pour l'intégrité de l'appareil auditif.

A.2 VOIE DE L'AUDITION

L'oreille humaine se décompose en trois parties principales : l'oreille externe localisée entre le pavillon et le tympan où l'information chemine *via* le conduit auditif, l'oreille moyenne définie entre le tympan et la fenêtre ovale, incluant les osselets (marteau, enclume et étrier) et l'oreille interne, formée de la cochlée (voir Figure 1 ci-dessous).

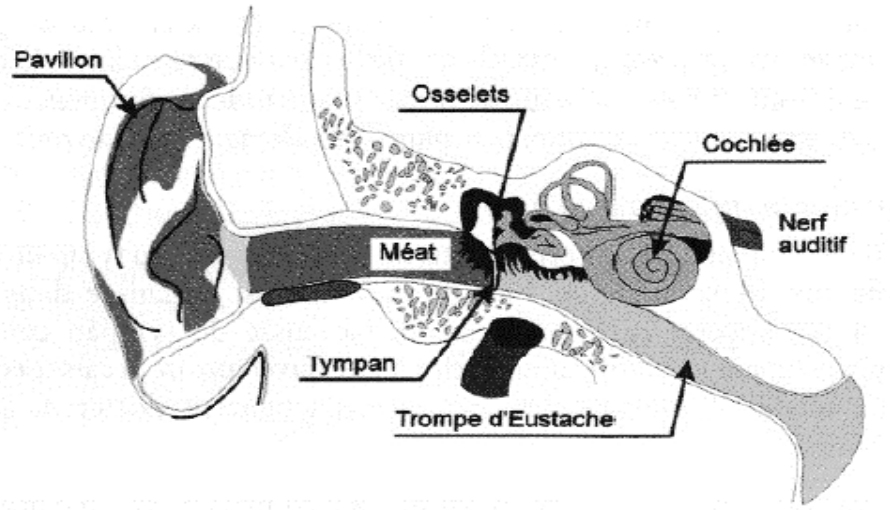


Fig. 1. Description anatomique de l'oreille (extrait de Laurent et Mathiot, 2007[27], p. 14).

La morphologie du pavillon et du conduit auditif (méat) permettent respectivement l'orientation (couplé au mouvement de tête) et l'amplification des sons. En effet, la pression acoustique est triplée entre son entrée dans le conduit auditif et son impact sur le tympan. La membrane tympanique reçoit les vibrations de l'air, acheminées *via* le canal auditif, et vibre à la même fréquence que l'onde sonore incidente. Par exemple, le tympan vibre lentement lors de sons graves, c'est-à-dire de faibles fréquences, et rapidement en présence de sons aigus, correspondant aux hautes fréquences. Cette conséquence vibratoire correspond au premier mécanisme d'une chaîne complexe à l'origine de l'audition.

Les vibrations du tympan vont ensuite être transmises à la fenêtre ovale, jonction entre l'oreille moyenne et la cochlée, grâce à la chaîne ossiculaire. Les mouvements de l'étrier vont

[27] Gérard, L. et Mathiot, D. (2007). *Techniques audiovisuelles et multimédia* : Tome 1 (2e éd.). Paris, France : Dunod.

permettre de relayer le signal acoustique de la fenêtre ovale à la cochlée. Le signal sera ensuite amplifié par les osselets pour éviter une perte trop importante entre l'oreille moyenne et le liquide cochléaire. L'oreille moyenne a aussi une fonction de protection de l'oreille interne, par un mouvement réflexe du muscle stapédien, en présence de sons de trop forte intensité (pression élevée). Le réflexe stapédien ou réflexe acoustique peut être comparé au réflexe de contraction pupillaire lors d'un signal lumineux trop intense. Cependant, en raison d'un temps de latence de 150 ms, ce réflexe est inefficace face aux sons impulsifs tels que des claquements de porte ou des bruits de tirs.

L'oreille interne ou la cochlée se présente comme un canal membraneux, en forme de spirale, emplie de liquide (endolymphe). La cochlée contient également des structures impliquées dans la sensation d'équilibre et de mouvements (canaux semi-circulaires). L'organe de Corti, qui repose sur la membrane basilaire, concentre les cellules réceptrices auditives ou cellules ciliées. Ces dernières sont fixées à la membrane basilaire et doivent leur nom à leur extension en forme de poil, baignant dans l'endolymphe. Les vibrations propagées dans le liquide cochléaire, en réaction aux mouvements de l'étrier, vont entraîner à leur tour la vibration de la membrane basilaire. Les cellules ciliées vont alors capter et amplifier les vibrations mécaniques transmises par les osselets et permettent la transduction des ondes sonores vibratoires en influx nerveux. Les cellules ciliées vont répondre préférentiellement à une fréquence donnée, selon l'endroit où elles se situent sur la membrane basilaire. En d'autres termes, le taux de décharge maximal de potentiels d'action sera rencontré en réaction à une fréquence caractéristique. Chaque région de la membrane basilaire correspond à une bande de fréquences caractéristique décroissante. Le signal transmis au nerf auditif par les cellules ciliées va donc contenir l'ensemble des différentes sources sonores analysé par bandes de fréquences. L'organisation spatiale (tonotopie) observée dans la cochlée va être retrouvée au niveau du nerf auditif, où une fréquence va correspondre à une fibre nerveuse, des relais sous-corticaux et des aires corticales auditives principalement situés au niveau du lobe temporal.

L'influx acoustique, issu des deux oreilles, va converger et être relayé, de la cochlée au cortex auditif (voie ascendante), grâce à quatre relais sous-corticaux : noyau cochléaire, olive supérieure, colliculus inférieur, corps genouillé médian. Le noyau cochléaire et l'olive permettent d'évaluer la position spatiale du stimulus sonore déterminé grâce, entre autres, aux différences d'intensité et aux délais d'arrivée du signal sonore entre les deux oreilles. Ces différences vont par exemple permettre de localiser la source latérale d'un son, si un son est plus intense ou arrive avant dans l'oreille gauche, le son sera localisé à gauche. Ensuite, les fibres convergent vers le colliculus inférieur pour relayer l'information jusqu'au cortex auditif *via* le corps genouillé médian (thalamus). Les réponses des neurones des noyaux sous-corticaux vont être sensibles à des caractéristiques de plus en plus complexes du signal entrant.

Le signal auditif entrant va être morcelé, au niveau du cortex auditif, en différents attributs (hauteur, sonie, timbre, position, *etc.*). Ces différents attributs seront ensuite préférentiellement traités, de manière parallèle, par les différentes aires du cortex auditif. Celui-ci se compose principalement du cortex auditif primaire (A1) et de plusieurs aires

associatives qui l'entourent (entre cinq ou six, Wallace, Johnston et Palmer, 2002[28]). L'organisation spatiale des relais sous-corticaux et de certaines aires du cortex auditif présente également une organisation tonotopique (Romani, Williamson et Kaufman, 1982[29]).

Le cortex auditif présente une organisation fonctionnelle hiérarchisée où A1 est davantage impliqué dans le traitement des attributs élémentaires (fréquence, intensité) tandis que les aires associatives seraient activées en présence de son de plus grande complexité (spectrale, tonale, spatiale, temporelle, *etc*.) (Wessinger *et al.*, 2001[30]). Ainsi, plus un son est complexe, plus l'activation du cortex auditif est étendue.

Par ailleurs, le traitement de l'information auditive pourrait se poursuivre au-delà du cortex auditif par l'emprunt d'une voie ventrale, impliquée dans les traitements de reconnaissance (voix familière), la voie du « What » et d'une voie dorsale, la voie du « Where », plutôt liée à la localisation d'un son (Kaas et Hackett, 1999[31]).

La localisation des sons et le traitement des sons en mouvement seraient effectués au niveau des cortex pariétaux et du cortex préfrontal tandis que la reconnaissance des sons serait réalisée au niveau du cortex frontal. *In fine*, le percept auditif fera appel à des composantes attentionnelles ou aux connaissances antérieures de l'auditeur.

B. SYSTEME VISUEL

Le système visuel humain (SVH) permet l'appréhension d'une grande quantité d'informations visuelles du monde extérieur comme la couleur, la taille, la forme, la texture, la distance ou encore la vitesse. Les paragraphes suivent décrivent succinctement les bases anatomophysiologiques et fonctionnelles du SVH dans l'intention de mieux comprendre la manière dont les informations visuelles sont captées et traitées. Le cheminement de la stimulation visuelle, tout d'abord sous sa forme physique jusqu'à sa traduction nerveuse, de l'œil aux différentes aires du cortex visuel, sera abordé.

---

28 Wallace, M. N., Johnston, P. W. et Palmer, A. R. (2002). Histochemical identification of cortical areas in the auditory region of the human brain. *Experimental brain research*, 143(4), 499-508.

29 Romani, G. L., Williamson, S. J. et Kaufman, L. (1982). *Tonotopic organization of the human auditory cortex. Science*, *216*, 1339-1340.

30 Wessinger, C., VanMeter, J., Tian, B., Van Lare, J., Pekar, J. et Rauschecker, J. (2001). Hierarchical organization of the human auditory cortex revealed by functional magnetic resonance imaging. *Journal of Cognitive Neuroscience*, *13*(1), 1-7.

31 Kaas, J. H. et Hackett, T. A. (1999). What and where processing in auditory cortex. *Nat. Neuroscience*, *2*(12), 1045-1047.

B.1 SIGNAL PHYSIQUE

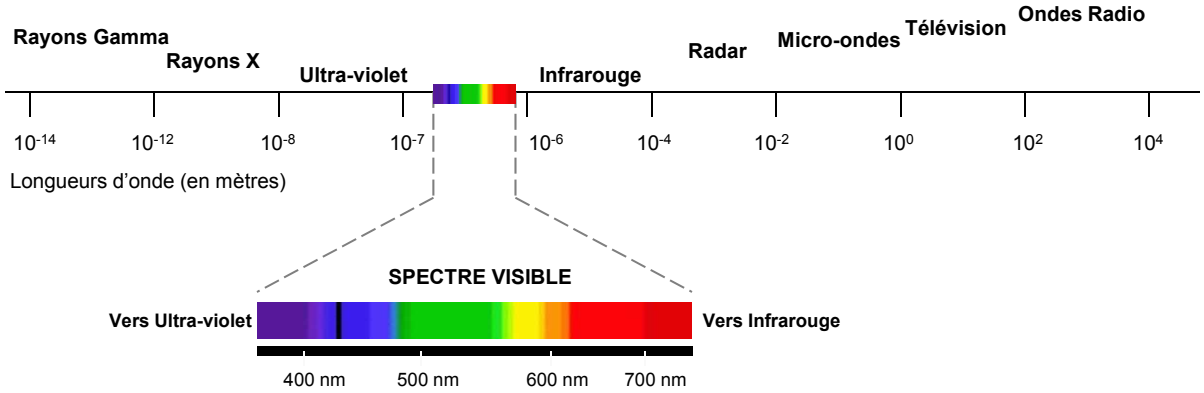


Fig. 2. Décomposition du spectre électromagnétique de 10^{-14} (ondes courtes invisibles) à 10^{4} (ondes longues invisibles). Le spectre visible présente une plage comprise entre 400 et 700 nanomètres (nm).

Les cellules photo-réceptrices de l'œil sont stimulées par les radiations électromagnétiques issues du spectre lumineux. Celles-ci se déplacent rapidement (~300 000 km/s) sous forme d'ondes dont la longueur peut varier de plusieurs kilomètres (10 km pour les ondes les plus longues comme les ondes radiophoniques) à seulement quelques nanomètres (10^{-14} mètres pour les ondes les plus courtes comme les rayons gamma). Sur ce vaste ensemble, les photorécepteurs rétiniens ne sont sensibles qu'à une plage étroite de longueurs d'ondes, définie entre quatre centre et sept cent nanomètres (les cellules rétiniennes n'étant plus capables de coder le stimulus en-deçà et au-delà de ces bornes), appelée lumière visible (voir fig. 2).

Les variations des longueurs d'ondes du spectre de lumière visible sont à l'origine de la vision de la couleur. Celle-ci résulte d'un processus de réflexion/absorption du spectre lumineux par un objet ou une surface donnée. Plusieurs scénarios peuvent survenir, en effet, le spectre lumineux peut : être totalement absorbé (aucune lumière n'est réfléchie : la surface est noire), traverser un objet (surface transparente : eau, verre) ou être totalement réfléchie (miroir ou surface blanche). En règle générale, les trois scénarios surviennent de manière concurrente. Par exemple, un objet va réfléchir une partie du spectre et en absorber une autre. C'est la cooccurrence de ces phénomènes qui va permettre la constitution du pigment d'une surface donnée. Par exemple, une surface de couleur verte correspondra à l'absorption des longueurs d'onde du bleu et du rouge et à la réflexion des longueurs d'onde associées à la couleur verte.

B.2 PROPRIETES OPTIQUES DE LA VISION : LA REFRINGENCE

La vision provient donc de la détection d'une partie des ondes lumineuses et la réflexion de toute ou partie de ce spectre visible à partir des objets de notre environnement permet la vision en couleur. L'œil est capable de recevoir cette image extérieure grâce au principe de réfringence.

Lorsque les ondes lumineuses sont réfléchies, à partir d'une source, elles se propagent dans toutes les directions et ce, à partir de chaque point de la source. Pour permettre la vision du monde extérieur, la convergence des ondes en un point unique est nécessaire. Cette focalisation sera à l'origine de la projection d'une image sur la rétine. Ainsi, la forme sphérique de l'œil n'est pas étrangère à ce phénomène d'incurvation, appelé réfraction, des rayons lumineux incidents.

Par exemple, lorsqu'un bâton est plongé dans l'eau, son reflet projette un angle non intuitif et donne l'illusion d'être déformer. Cette déformation s'explique par la densité du milieu. Lorsque le signal lumineux rencontre une surface de densité différente à celle de l'air, comme de l'eau, sa direction dévie selon un angle qui va dépendre d'une part de la densité et d'autre part, de l'angle d'impact entre l'onde et la surface.

Les surfaces courbes de l'œil (cornée, cristallin, corps vitré, humeur aqueuse) utilisent ce même mécanisme pour incurver et rassembler l'ensemble des ondes lumineuses en un seul point sur la rétine. Avant d'atteindre les cellules photo-réceptrices de la rétine, le signal lumineux doit en effet traverser l'ensemble de ces couches dites réfringentes. C'est grâce à ce mécanisme de réfraction des ondes lumineuses qu'une image nette pourra être focalisée au niveau de la rétine. L'information visuelle ainsi réfractée sur la rétine résulte en une image inversée, c'est-à-dire que chaque hémi-champ (champ visuel gauche et champ visuel droit) se projette sur la demi-rétine qui lui est opposée (nasale ou temporale).

C'est ensuite la transduction de cette stimulation en un influx nerveux qui va permettre la perception du monde extérieur. Les bases anatomophysiologiques et les fonctions de l'œil et du SVH dans son ensemble sont décrites dans les paragraphes suivants.

B.3 VOIE DE LA VISION

La principale fonction de l'œil est de capter les rayons lumineux incidents, notamment grâce à sa composante optique et de permettre la focalisation des images issues du monde extérieur sur les cellules réceptrices présentes dans la rétine (principe de réfringence : réfraction des ondes pour convergence en un point unique focalisé sur la rétine). Une seconde fonction réside dans sa capacité à transformer cette stimulation visuelle en influx nerveux grâce à la transduction de l'information électromagnétique en potentiels d'actions. L'information nerveuse sera ensuite transmise aux différentes aires visuelles *via* les nerfs optiques.

B.3.1 ANATOMIE DE L'ŒIL

D'un point de vue général, l'œil se présente comme un globe rempli de liquide et protégé de trois parois distinctes divisibles en deux chambres : la chambre antérieure (entre l'iris et la cornée, remplie d'humeur aqueuse pour le maintien de la pression intra-oculaire) et la chambre postérieure (entre le cristallin et la rétine, remplie d'humeur vitrée soit 90% du volume de l'œil, pour le maintien de la rigidité de l'œil). La Figure 3 ci-dessous présente l'anatomie générale de l'œil.

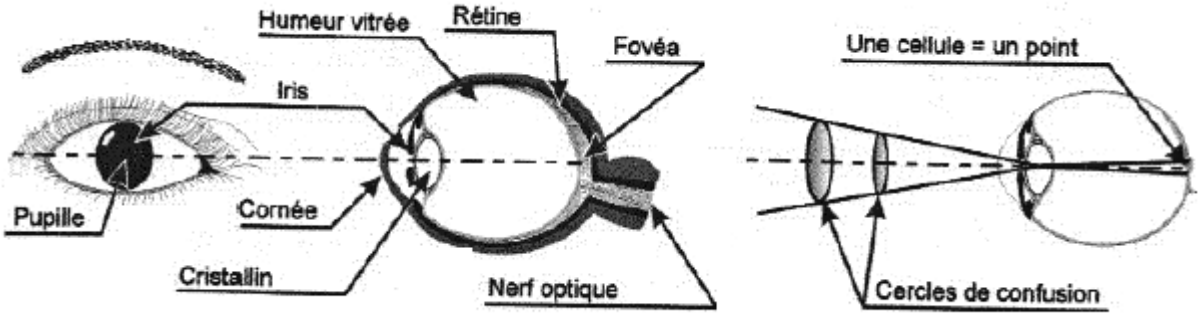


Fig. 3. Description anatomique de l'œil (extrait de Laurent et Mathiot, 2007, p.23).

L'enveloppe externe de l'œil, la sclère (ou blanc de l'œil), protège la totalité du globe oculaire et constitue le point d'accroche des muscles externes. Ceux-ci permettent la mobilité des globes oculaires à l'intérieur de leurs orbites. Sur la partie antérieure de l'œil, la sclère se confond avec la cornée, seul tissu transparent et avasculaire du corps humain, permettant le passage de la lumière.

L'uvée comprenant la choroïde (chambre postérieure), le corps ciliaire et l'iris (chambre antérieure) constitue la paroi intermédiaire de l'œil. La choroïde, tissu fortement vascularisé (artères, veines, nerfs) présente une fonction nutritive (irrigation de la rétine) et protectrice contre la diffusion de la lumière sur le fond l'œil (garantit la netteté de l'image par le maintien de l'intérieur de l'œil en chambre noire). Le muscle ciliaire contrôle la courbure du cristallin (lentille suspendue au corps ciliaire, derrière l'iris), et donc sa réfringence, permettant ainsi l'accommodation pour la vue de près ou de loin. Plus précisément, le muscle ciliaire se contracte pour l'accommodation de près (forme arrondie du cristallin) et se relâche pour la vision de loin (aplatissement du cristallin). Cet effort de déformation permet de maintenir une image nette de l'environnement indépendamment d'une certaine distance. L'iris (dont le pigment détermine la couleur des yeux) fait fonction de diaphragme en contrôlant la taille de la pupille (ouverture permettant l'entrée de la lumière) plus ou moins grande selon le niveau de luminosité (2 à 8 mm de diamètre, Delorme et Flückiger, 2003, p.76 [32]).

La rétine est la membrane nerveuse la plus interne de l'œil. Cette membrane formée de plusieurs couches cellulaires contient différentes classes de neurones. La rétine est le premier niveau traitement de l'information visuelle et ses neurones constituent les premières couches du SVH. Ses cellules nerveuses, absorbant l'énergie lumineuse, se divisent en deux types de photorécepteurs fonctionnellement et structurellement distincts : les cônes (5 à 6 millions, de forme conique) et les bâtonnets (100 à 130 millions, de forme cylindrique). Les premiers se concentrent principalement au centre de la rétine dans une zone appelée fovéa (ou encore macula ou tâche jaune). La fovéa est la région centrale du champ visuel, elle se situe face à la pupille (axe visuel) et correspond à l'endroit de la rétine où l'acuité visuelle est la plus élevée.

[32] Delorme, A., & Flückiger, M. (2003). *Perception et réalité: Introduction à la psychologie des perceptions*. Bruxelles, Belgique : De Boeck Supérieur.

La densité des cônes diminue drastiquement en allant vers la périphérie de cette zone. Une particularité des cônes est de n'être fonctionnels qu'en présence de lumière. Ils se définissent notamment par leur sensibilité aux variations de longueurs d'ondes du spectre lumineux et leur principale fonction est d'être impliquée dans la vision chromatique. Celle-ci s'effectue grâce à trois types de cônes spécifiquement sensibles aux variations de longueurs d'ondes du spectre lumineux (ondes longues/rouge, ondes moyennes/vert, ondes courtes/ bleu). Les bâtonnets sont totalement absents de la fovéa mais se distribuent sur tout le reste de la rétine. Ceux-ci sont impliqués dans la vision scotopique (vision nocturne, 10^{-6} à 10^{-2} candelas/m^2) et ne permettent pas de coder les différentes longueurs d'onde (aucune information de couleur).

Les cônes et les bâtonnets permettent la transformation du stimulus lumineux en influx nerveux, c'est-à-dire que le signal physique est converti en signal électrique. Celui-ci sera ensuite transmis aux différentes aires cérébrales *via* le nerf optique.

B.3.2 APRÈS LA RÉTINE

Après avoir traversé les différentes parois de l'œil, le signal lumineux est transformé en messages nerveux grâce aux différentes cellules rétiniennes. Le signal nerveux résultant est ensuite transporté *via* les nerfs optiques jusqu'à un premier relai, le chiasma optique (zone où se rejoignent les fibres provenant des deux rétines). L'information est ensuite acheminée *via* les tractus optiques (directs ou croisés) vers les différentes couches cellulaires des corps genouillés latéraux (CGL) du thalamus, à l'exception de quelques axones cheminant vers les colliculi supérieurs (environ 1% des fibres). Une fonction majeure du corps genouillé latéral serait son implication dans la sélection attentionnelle de la modalité visuelle. Les radiations optiques (axones issus des cellules du CGL) vont se projeter vers l'aire visuelle primaire du cortex. Chaque hémi-champ visuel sera traité par le cortex visuel de l'hémisphère controlatéral.

L'information visuelle est ensuite parcellisée en différentes propriétés pour être traitée distinctement par différentes parties du système visuel. La détection des contours, des contrastes, de la couleur, du mouvement et de son orientation, de la forme ou encore les aspects stéréoscopiques de la vision sont ensuite traités par les différents aires du cortex visuel. L'information visuelle est donc traitée de manière parcellaire (forme, couleur, mouvement, *etc.*) et parallèle dans les différentes aires corticales, l'intégration synchrone des informations issues des différentes aires permettra d'aboutir à la perception visuelle finale. La reconnaissance et l'interprétation des informations visuelles est ensuite réalisée par la mise en relation avec les connaissances antérieures, stockées en mémoire ou encore des facteurs attentionnels et/ou émotionnels.

B.4 MOUVEMENTS OCULAIRES

La vision de scènes ne peut s'effectuer sans les mouvements oculaires. Ceux-ci vont permettre de positionner la fovéa, où l'acuité est maximale, sur la zone d'attention. Ainsi, la vision ne peut s'envisager sans prendre en compte les mouvements effectués par les yeux.

Les mouvements de la tête et les mouvements oculaires, plus rapides et précis, permettent de maintenir le point de fixation (zone importante de l'image) sur la rétine et plus particulièrement sur la fovéa où l'acuité est maximale. Par exemple, l'activité de lecture implique un déplacement constant de l'œil pour maintenir en zone fovéale les lettres devant être lues (empan visuel de 3 lettres de chaque côté du point de fixation). Ainsi pour permettre la vision (fovéale ou périphérique), les yeux ne sont jamais immobiles.

Le globe oculaire est mobilisé grâce à la coordination de six muscles extrinsèques (fig. 4 ci-dessous). Ce corpus musculaire permet des mouvements selon trois axes : transversal (haut/bas), antéro-postérieur (intorsion/extorsion) et vertical (temporal/nasal). Les différents mouvements oculaires peuvent être regroupés selon leurs caractéristiques physiques (lents ou rapides), réflexes (réflexes ou volontaires) ou directionnels (conjonctifs ou disjonctifs). Le déclenchement de ces mouvements permet de maintenir une cible fixe (saccade ou vergence) ou mobile (poursuite oculaire) sur la fovéa. Les mouvements oculaires sont aussi parfois des réactions réflexes aux mouvements de l'observateur ou de l'environnement (réflexe optocinétique ou vestibulo-oculaire). Les paragraphes suivant s'attacheront à décrire, du point de vue fonctionnel, les différents mouvements oculaires.

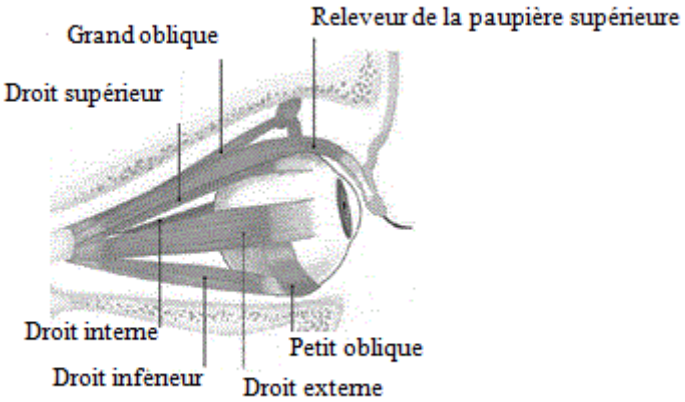


Fig. 4. Les différents muscles moteurs du globe oculaire (d'après le Lacombe[33], 2009).

[33] Voir REFERENCES

B.4.1 VISION BINOCULAIRE

Lorsque les deux yeux fixent un objet donné, une image se fixe alors sur la rétine de chaque œil. Cependant, en raison de la distance interoculaire (en moyenne 6 cm), la distance de l'image reflétée sur l'œil gauche n'est pas tout à fait identique à celle reflétée par l'œil droit. Le traitement de cette disparité rétinienne par les neurones binoculaires du cortex visuel primaire est à l'origine de la perception du relief (stéréoscopie). La stéréoscopie est le mécanisme permettant la vision tridimensionnelle grâce aux traitements des informations de disparités rétiniennes. Diverses techniques utilisent cette information de disparité pour reproduire la vision stéréoscopique sur différents types de supports, photographiques ou cinématographiques par exemple. L'essor du cinéma 3D est aujourd'hui un exemple parlant de l'utilisation technologique du phénomène de stéréoscopie pour permettre au spectateur de percevoir artificiellement les informations tridimensionnelles.

Ainsi, le relief et la profondeur ne sont pas présents, en tant que tel, au niveau du signal sensoriel rétinien mais seront reconstruits à partir d'un traitement cognitif capable de traiter ce type d'afférences binoculaires.

B.4.2 VERGENCE

Sekuler, R. et Blake (1990[34]) ont décrit deux types de mouvements : les mouvements conjonctifs et les mouvements de vergence. Les premiers concernent les déplacements oculaires conjoints (saccades ou poursuites), c'est-à-dire que les déplacements sont effectués dans la même direction tandis que les seconds sont disjonctifs, c'est-à-dire que les déplacements s'opèrent dans des directions opposées. Selon que cette vergence soit nasale (proximité) ou temporale (distance), on parle respectivement de convergence ou de divergence. Les mouvements de vergence ont pour fonction d'aligner la fovéa avec des cibles selon différentes distances. En cas de convergence, une mise au point de la cible est réalisée grâce à une modification de la courbure du cristallin. Par ailleurs, une constriction pupillaire s'ajoute pour améliorer la netteté de l'image en augmentant la profondeur de champ. Lorsque que les yeux convergent, ils forment un angle d'autant plus grand que l'objet fixé sera proche. L'angle de convergence permet de déterminer la distance d'un objet jusqu'à un maximum de six mètres. Au-delà de cette distance, l'angle binoculaire est trop petit pour être exploitable (Sekuler, R. et Blake).

Les mouvements de vergence vont permettre l'obtention d'une image sur la fovéa. Cependant, le « maintien fovéal » de cette image va être obtenu selon deux types de mouvements basiques : les mouvements rapides (saccades) et les mouvements lents (poursuite).

B.4.3 POURSUITE OCULAIRE

Le comportement de poursuite oculaire (systèmes optocinétique) correspond à des mouvements lents initiés en cas de poursuite de cible traversant le champ visuel (balle de

[34] VOIR REFERENCES

tennis, personnage, formule 1, *etc*.). Ces mouvements permettent de poursuivre une cible mobile pour en maintenir l'image sur la fovéa. Face à des scènes visuelles impliquant la poursuite d'élément de grande vitesse (passage d'un train par exemple), les yeux suivent la cible jusqu'à leur déviation maximale (limite de l'orbite) pour ensuite, grâce à une saccade réflexe en sens inverse, reprendre un mouvement lent de poursuite. Ce mouvement réflexe est appelé le nystagmus ou réflexe optocinétique.

Un second type de réflexe, le réflexe vestibulo-oculaire (RVO), permet de garder une image fixe sur la fovéa durant les mouvements de la tête. Le RVO survient avec un temps de latence relativement court (environ 14 ms) pendant les déplacements de la tête et permet donc le maintien d'une image stable.

B.4.4 FIXATIONS

Les fixations correspondent à des périodes, comprises entre 200 et 600 ms, relativement stables, pendant lesquelles un objet peut être vu. Les fixations permettent la projection de la zone d'intérêt sur la fovéa et l'analyse conséquente avec un maximum de discrimination spatiale (Yarbus, 1967[35]). La fixation peut donc se définir comme l'activité des yeux lorsque ceux-ci ne bougent pas. Il est généralement accepté que les traitements visuels et cognitifs sont réalisés durant ces périodes de fixations (Just et Carpenter, 1984). L'information de la scène analysée serait extraite durant ces fixations (Rayner, 1998[36] ; Wedel et Pieters, 2000[37]). Le nombre de fixations, plus que leurs durées, refléterait le degré d'information extrait d'un contenu (Wedel et Pieters, 2000). Cependant, l'œil n'est pas totalement immobile durant les fixations. En effet, des mouvements très rapides (< 1°/s) et de très petite amplitude (20 à 40 secondes d'angle) permettent d'éviter à l'image de s'évanouir (Kowler, 1990[38]; Yarbus, 1967).

B.4.5 SACCADES

La saccade peut se définir comme tout mouvement oculaire rapide destiné à modifier la direction du regard. En effet, les mouvements de saccades permettent d'amener l'œil (plus précisément la fovéa) rapidement (vélocité angulaire jusqu'à 800°/s selon Berthoz et Petit, 1996[36]) et précisément sur une nouvelle zone d'intérêt. Les saccades correspondraient aux mouvements les plus rapides pouvant être réalisés par le corps humain (Stern R. *et al.*, 2001[36]). La durée (entre 50 et 150ms) et la vitesse d'une saccade vont dépendre de son

---

35 Yarbus, A. L. (1967). *Eye movements and vision*. New York, NY : Plenum Press.

36 Voir REFERENCES

37 Wedel, M. et Pieters, R. (2000). Eye fixations on advertisements and memory for brands: A model and findings. *Marketing science*, *19*(4), 297-312.

38 Kowler, E. (1990). The role of visual and cognitive processes in the control of eye movement. Dans E. Kowler (dir.), *Eye movements and their role in visual and cogntive processes* (Vol. 4, p. 1-70). Amsterdam, Pays-Bas: Elsevier.

amplitude, c'est-à-dire la distance à parcourir jusqu'au nouveau point de fixation. Plus l'amplitude augmente, plus la saccade sera rapide et longue (Boghen, Troost, Daroff, Dell'Osso et Birkett, 1974[39]). Les saccades surviennent très fréquemment (> 1763 000 fois/jour).

Malgré la vitesse de ces déplacements, la vision reste stable. Cette modification de la position du regard permet la projection d'une nouvelle image sur la fovéa des deux rétines. La vitesse et le nombre de saccades est assujetti aux variabilités inter-individuelles telles que l'âge et intra-individuelle (fatigue, environnement, *etc.*).

Les saccades permettent notamment d'explorer une scène visuelle par l'alternance d'instants de fixation (assimilation de l'information) et de déplacements qui portent le regard jusqu'à un nouvel espace de fixation. Ces mouvements balistiques (pas de redirection possible une fois le mouvement enclenché) peuvent être d'amplitude variable selon que l'activité implique une fenêtre d'exploration large (balayage d'une pièce) ou réduite (lecture). Le temps d'enclenchement d'une saccade varie généralement de 90 à 150 ms.

Leur déclenchement peut être volontaire (exploration d'une scène) ou réflexe (saccades extrêmement rapides d'environ 1 degré d'arc) en cas de fixation prolongée (nystagmus optocinétique ou réflexe vestibulo-oculaire) ou de l'arrivée soudaine d'un nouveau stimulus indépendamment de sa modalité (Leigh et Zee, 2006[40]).

Lors d'une exploration de scène, les déplacements oculaires ne se réalisent pas de manière hasardeuse mais suivent une stratégie déterminée par une hiérarchie informationnelle. Par exemple, les yeux le nez, la bouche et les oreilles seront les zones préférentiellement fixées sur un visage car elles correspondent aux zones les plus riches, c'est-à-dire regroupant les informations les plus pertinentes pour la compréhension de l'image ou de la scène (Yarbus, 1967). Les saccades jouent un rôle important dans la construction des représentations visuelles de l'environnement (Rayner, 1998) et reflètent également les mouvements de l'attention visuelle (Remington, 1980[41]). Ainsi, les mouvements de saccades influencent la perception qui en résulte. Les mouvements de saccades sont contrôlés notamment par le colliculus supérieur (Koch, 2004[41]). En effet, ce dernier reçoit les afférences directes de la rétine (environ 1% des axones).

---

39 Boghen, D., Troost, B., Daroff, R., Dell'Osso, L. et Birkett, J. (1974). Velocity characteristics of normal human saccades. *Investigative Ophthalmology and Visual Science, 13*(8), 619-623.

40 Leigh R. J. et Zee, D. S. (2006). *The neurology of eye movements* (4e éd.). New York, NY : Oxford University Press..

41 Voir REFERENCES

ANNEXE 3-A

Approche phasique pour les mesures d'AED

A. LES RED

Lors de la présentation d'un stimulus, des variations de l'AED peuvent être observées. Les RED traduisent l'activité sympathique en réaction à une stimulation. La phase ascendante de la réponse représente la production de sueur (stimulation des glandes eccrines), le début de la phase de récupération annonce l'arrêt de la stimulation des glandes (arrêt de la sécrétion de sueur) et le début d'un phénomène d'évaporation ou d'absorption de la sueur. Les RED se produisent entre une et trois secondes après la présentation d'un stimulus (Venables et Mitchell, 1996[42]) avec une amplitude généralement comprise entre 0,2 et 1 µS. Elles reflètent une activation sympathique à l'origine de l'excitation des glandes sudoripares eccrines provoquant une micro-sudation. Le temps de latence entre l'initiation de la RED et la survenue du pic est généralement de une à trois secondes, il faut ensuite compter environ dix secondes pour que l'AED retrouve son niveau basal (Dawson *et al.*, 2007[43]). La RED débute par une augmentation rapide de la conductance jusqu'au maximum local (pic) puis une phase de récupération, plus lente, survient. Cependant, des micro-variations rapides (proche du maxima) peuvent être observées avant la phase de récupération (Vernet-Maury, Robin et Dittmar, 1995[44]). Ces différents paramètres sont illustrés par la Figure 1 ci-après.

42 Venables, P.H. et Mitchell, D. (1996). The effects of age, sex and time of testing on skin conductance activity. *Biological Psychology, 43*(2), 87-101.

43 Voir REFERENCES

44 Vernet-Maury, E., Robin, O. et Dittmar, A. (1995). The ohmic perturbation duration, an original temporal index to quantify electrodermal responses. *Behavioural brain research, 67*(1), 103-107.

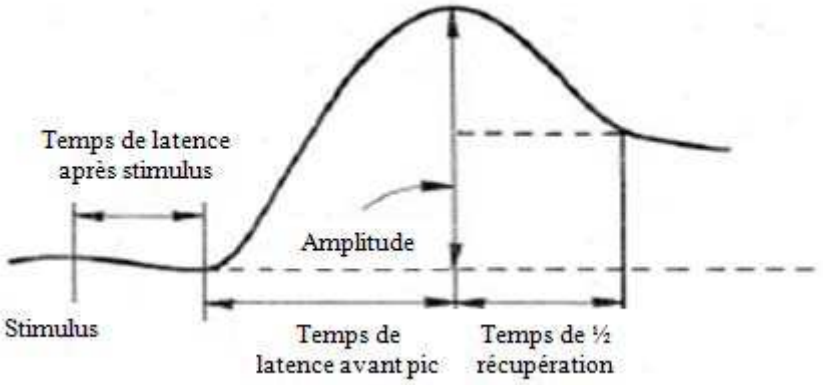


Fig. 1. Représentation graphique des principaux paramètres d'une RED (adapté de Dawson *et al.*, 2007)

Par ailleurs, selon Dawson *et al.* (2007), un phénomène d'habituation peut être observé après la présentation d'un certain nombre de stimuli (entre 2 et 8). Cette habituation se traduit par une disparition des RED en réaction aux stimuli présentés. Toutefois ce phénomène d'habituation serait fonction de l'individu, certains y étant plus sensibles (individu labile) et d'autres plus résistant (individu stabiles).

A.1 TRAITEMENT DES RED

Comme indiqué précédemment, les RED surviennent entre 1 et 3 secondes après la présentation d'un stimulus. Venables et Christie (1980[45]) proposent donc d'étudier les RED comprises dans cet intervalle de latence. Toute RED débutée dans cette fenêtre temporelle pourra alors être considérée comme conséquente au stimulus présenté. Ainsi, l'absence de RED lors de l'investigation de cette fenêtre reflète une non-réactivité de l'AED au stimulus précédemment présenté. En d'autres termes, l'AED peut être étudié au traves de la présence ou non de RED post-stimulus.

Plusieurs indicateurs permettent ensuite de quantifier les RED présentes. L'indicateur le plus fréquemment utilisé pour décrire les RED correspond à l'étude de l'amplitude. Cet indicateur correspond à la différence entre le début de la RED et son maximum local, l'amplitude minimum à partir de laquelle une réponse peut être considérée est fixée à 0,05 µS (Gruzelier et Venables, 1972[46]). Du point de vue physiologique, l'amplitude est liée à l'activation sympathique.

Un second indicateur correspond à l'étude de la durée. En effet, les RED étant des phénomènes transitoires de l'AED, elles ont une durée limitée dans le temps. Du point de vue physiologique, plus l'activation sympathique serait intense et/ou soutenue dans le temps plus

---

[45] Voir REFERENCES

[46] Gruzelier, J. et Venables, P.H. (1972). Skin conductance orienting activity in a heterogeneous sample of schizophrenics: Possible evidence of limbic dysfunction. *Journal of Nervous and Mental Disease, 155*(4), 277-287.

les glandes sudoripares eccrines seraient stimulée et par conséquent, plus la RED serait longue (Clarion[47], 2009, p.112). En effet, plus de sueur sera produite plus le temps d'évaporation ou de réabsorption sera long. Une RED se décompose en deux temps : l'ascension (début de la pente à son maximum) et la récupération (maximum au retour à la valeur basale). Ces deux phases permettent de caractériser la RED en matière de durée. La durée moyenne d'ascension est comprise entre 2 et 3 secondes pour une plage comprise entre 0,5 et 5 secondes (Boucsein[47], 2012). La phase de récupération n'est pas étudiée dans sa totalité en raison de la difficulté d'un retour à la valeur basale de départ (variations toniques de l'AED). Pour pallier cette difficulté, le temps de demi-récupération est alors utilisé comme indicateur de la phase descendante. Il s'agit du temps mis pour récupérer la moitié de l'amplitude de la RED. En règle générale, les durées de demi-récupérations sont comprises entre 3 et 5 secondes (Boucsein, 2012).

L'étude de la pente de l'ascension est un autre paramètre permettant de caractériser une RED. Celle-ci peut-être calculée entre le début et la fin de la RED ou entre le début et le maximum de la pente (pente moyenne de l'ascension). Enfin, d'autres paramètres comme l'aire de la RED permettent de quantifier une RED mais ne seront pas détaillés dans ce document.

B. RED-NS

Lors de l'apparition de stimuli, des réactions ou réponses électrodermales (RED) peuvent être observées au sein de l'AED. Ces réactions transitoires, de durées limitées, sont également observées en l'absence d'évènement particulier. Ces réactions non spécifiques (RED-NS, Boucsein, 2012) surviennent généralement toutes les une à trois minutes en situation de repos (Dawson *et al.*, 2007). En dehors de ces évènements réguliers, les RED-NS présente une variabilité intra-individuelle. Nikula (1991, cité par Clarion, 2009), a montré que la fréquence des RED-NS augmentait lorsque des émotions négatives, des discours internes, des états d'activation ou de préoccupation étaient subjectivement évalués. Ce résultat indique que les RED-NS sont soumises aux influences des différents processus cognitifs mis en jeu.

En plus de ces variations intra-individuelles, les différents indicateurs de l'AED, NED, RED-NS et RED, sont sujettes à une forte variabilité inter-individuelle.

[47] Voir REFERENCES

ANNEXE 6-A

Conditions de visualisation et d'écoute - Expérimentation A

Correspondances entre les conditions de visualisation et d'écoute recommandées par la norme UIT-T P.911 et celles de l'expérimentation A. Les dimension de la pièce sont exprimées de la manière suivante : (longueur)×(Largeur)×(Hauteur).

PARAMETRES	REGLAGES P.911	REGLAGES EXPE.A
Dimensions de la pièce	l×L×H	520×370×285 cm
Distance de visualisation	De 1 à 8 H	Respecté (6 H)
Luminance de l'écran (valeur de crête)	de 100 à 200 cd/m2	Respecté
Rapport de luminance d'écran inactif à luminance de crête	≤ 0,05	Respecté
Rapport de luminance de l'écran au niveau de crête du blanc (lors de l'affichage d'un niveau de noir total dans une salle complètement obscure)	≤ 0,1	Respecté
Rapport entre luminance de l'arrière-fond du moniteur d'image à la valeur de crête de la luminance d'image	≤ 0,2	Respecté
Chromaticité de l'arrière-fond	D65	Non respecté (x= 0.4855 ; y =0.4078)
Eclairement lumineux d'ambiance de la salle	≤ 20 lx	Respecté
Niveau de bruit de fond	≤ 30 dBA	Respecté
Niveau d'écoute	~ 80 dBA	Respecté
Durée de réverbération	< 500 ms, $\forall f > 150$ Hz	Respecté

ANNEXE 6-B

Questionnaire d'évaluation - Expérimentation A

Présentation des échelles utilisées pour l'évaluation de la qualité audiovisuelle, vidéo et audio. L'évaluation était réalisée après la visualisation de chaque séquence de test. Les participants avaient pour consignes de cercler le chiffre correspondant à leur opinion.

Evaluation Extrait *A* : Visualisation 1

1) Comment évaluez-vous la qualité **audiovisuelle globale** ? (cerclez un numéro)

9 Excellent
8
7 Bon
6
5 Satisfaisant
4
3 Médiocre
2
1 Mauvais

2) Comment évaluez-vous la qualité **vidéo** de l'extrait? (cerclez un numéro)

9 Excellent
8
7 Bon
6
5 Satisfaisant
4
3 Médiocre
2
1 Mauvais

3) Comment évaluez-vous la qualité **audio** de l'extrait? (cerclez un numéro)

9 Excellent
8
7 Bon
6
5 Satisfaisant
4
3 Médiocre
2
1 Mauvais

ANNEXE 6-C

Consignes - Expérimentation A

Bonjour et merci de votre participation.

Dans cette expérience, vous allez visualiser 3 extraits audiovisuels différents : documentaire *boxe* (x3), match de *tennis* (x3) et *opéra* (x3). Ces 3 extraits vous seront présentés successivement. Chaque extrait sera vu trois fois à la suite.
Au total 9 extraits vous seront donc présentés.

A chacun de ces 9 extraits correspondra un questionnaire présentant 3 échelles distinctes. Vous aurez, en tout et pour tout, 9 questionnaires à remplir soit un total de 27 échelles.

Chaque fois que vous verrez et entendrez un extrait, vous devrez en juger la qualité en cerclant, pour chacune des 3 échelles présentées, l'un des 9 niveaux reflétant votre opinion :

9	Excellent
8	
7	Bon
6	
5	Satisfaisant
4	
3	Médiocre
2	
1	Mauvais

Chaque échelle correspondra à une évaluation différente :

1. Evaluation de la qualité audiovisuelle globale, **audio et vidéo**, combinée.
2. Evaluation de la qualité **vidéo**
3. Evaluation de la qualité **audio**

Vous disposerez de quelques minutes entre chaque extrait pour remplir le questionnaire correspondant à l'extrait venant d'être visualisé/entendu. Ce laps de temps sera signalé par un écran noir.
Observez et écoutez avec attention l'ensemble de l'extrait avant d'exprimer votre jugement.

Consignes de posture :

Il vous sera demandé durant le test de :

- choisir une posture confortable et de la maintenir autant que possible tout au long du test,
- ne pas positionner votre main libre devant votre bouche,
- garder la main choisie pour l'installation des capteurs, la plus immobile possible.

ANNEXE 6-D

Tableaux de résultats (ANOVAs) - Expérimentation A

Résultats des ANOVAs réalisées à partir des jeux de données JD-5min et JD-30s considérant respectivement les variables indépendantes « Contenu » et « Qualité » ; « Contenu », « Qualité » et « Périodes » et les variables dépendantes (VD) physiologiques et oculaires.

Résultats issus de JD-5min

JD-5min	VD	Somme des carrés	ddl effet	ddl erreur	Moyenne des carrés	F	*p*
Contenu	**AED*n***	2,28	2	56	1,14	7,43	<0,01
Qualité	**AED*n***	0,29	2	56	0,15	5,22	<0,01
Qualité*contenu	**AED*n***	0,29	4	112	0,07	2,93	<0,05
Contenu	**FC*n***	0,01	2	56	0,01	2,83	0,07
Qualité	**FC*n***	0,00	2	56	0,00	1,68	0,20
Qualité*contenu	**FC*n***	0,00	4	112	0,00	1,80	0,13
Contenu	**TCP*n***	0,00	2	56	0,00	1,13	0,33
Qualité	**TCP*n***	0,00	2	56	0,00	0,83	0,44
Qualité*contenu	**TCP*n***	0,00	4	112	0,00	3,19	<0,05
Contenu	**VSP*n***	0,00	2	56	0,00	10,95	<0,001
Qualité	**VSP*n***	0,00	2	56	0,00	2,25	0,12
Qualité*contenu	**VSP*n***	0,00	4	112	0,00	1,41	0,23
Contenu	**DP**	0,00	2	46	0,00	30,58	<0,001
Qualité	**DP**	0,00	2	46	0,00	0,30	0,74
Qualité*contenu	**DP**	0,00	4	92	0,00	0,32	0,86
Contenu	**EBdur**	0,01	2	46	0,00	3,19	0,05
Qualité	**EBdur**	0,00	2	46	0,00	0,92	0,41
Qualité*contenu	**EBdur**	0,01	4	92	0,002	3,58	<0,01
Contenu	**Ebfreq**	0,60	2	46	0,30	2,36	0,11
Qualité	**Ebfreq**	0,08	2	46	0,04	0,21	0,81
Qualité*contenu	**Ebfreq**	0,58	4	92	0,14	0,77	0,55
Contenu	**PERCLOS**	0,00	2	46	0,00	0,52	0,60
Qualité	**PERCLOS**	0,00	2	46	0,00	2,26	0,12
Qualité*contenu	**PERCLOS**	0,00	4	92	0,00	0,33	0,86

Résultats issus de JD-30s

JD-30s	VD	Somme des carrés	Ddl effet	Ddl erreur	Moyenne des carrés	F	*p*
Contenu	**AED*n***	22,77	2	56	11,39	7,43	<0,01
Qualité	**AED*n***	2,94	2	56	1,47	5,22	<0,01
Périodes	**AED*n***	12,51	9	252	1,39	21,21	<0,001
Contenu	**FC*n***	0,13	2	56	0,07	2,83	0,067
Qualité	**FC*n***	0,05	2	56	0,02	1,68	0,195
Périodes	**FC*n***	0,12	9	252	0,01	5,00	<0,001
Contenu	**TCP*n***	0,04	2	56	0,02	1,13	0,329
Qualité	**TCP*n***	0,00	2	56	0,00	0,83	0,443
Périodes	**TCP*n***	0,00	9	252	0,00	2,47	<0,05
Contenu	**VSP*n***	0,00	2	56	0,00	10,23	<0,001
Qualité	**VSP*n***	0,00	2	56	0,00	1,76	0,181
Périodes	**VSP*n***	0,00	9	252	0,00	1,11	0,358
Contenu	**DP**	0,00	2	46	0,00	30,58	<0,001
Qualité	**DP**	0,00	2	46	0,00	0,30	0,744
Périodes	**DP**	0,00	9	207	0,00	4,29	<0,001
Contenu	**EBdur**	0,06	2	46	0,03	3,19	0,050
Qualité	**EBdur**	0,02	2	46	0,01	0,92	0,407
Périodes	**EBdur**	0,08	9	207	0,01	6,18	<0,001
Contenu	**Ebfreq**	2,44	2	46	1,22	1,65	0,203
Qualité	**Ebfreq**	0,03	2	46	0,02	0,02	0,981
Périodes	**Ebfreq**	0,33	9	207	0,04	1,08	0,382
Contenu	**PERCLOS**	0,00	2	46	0,00	0,39	0,681
Qualité	**PERCLOS**	0,01	2	46	0,00	2,56	0,088
Périodes	**PERCLOS**	0,18	9	207	0,02	7,48	<0,001

ANNEXE 7-A

Descripteurs pour caractérisation experte

Présentation de la liste des trente deux descripteurs utilisés par l'expert pour décrire le corpus de contenus audiovisuels.
Contenu : Danse/Documentaire/Opéra/Sport/Théâtre
Séquence n° : …
Intervalles (min,ss) et durée : …

SEMANTIQUE

Sémantique Générale			
1. Modalité Dominante	audio	vidéo	audio/Vidéo
2. Mouvement	avec	sans (statique)	
3. Info. textuelles	oui	non	

Dynamique			
4. Dynamique contenu (rythme)	faible	modérée	forte
5. Dynamique caméra	faible	modérée	forte

Paramètres scénaristiques
6. Extérieur - Intérieur
7. Jour – nuit
8. Clair - sombre
9. Visuels - dialogues
10. Intime – collectif - publique
11. Nombre personnages :…. dont Homme : … et Femme : ….
12. Inaction - action

Segment Audio			
13.Expressions sonores	paroles	musique	bruit
14.Paroles	Dialogue-monologue	commentaires	chant

Segment AV			
15.Rapport AV	son *in* (diégétique)	son hors-champ	son *off* (extra-diégétique)

TECHNIQUE

Technique générale			
16.Niveau de détail	faible	modéré	fort
17.Température de couleur	chaude (orangé)	jour (blanc)	froide (bleu)
18.Luminosité	faible	modérée	forte

Caméra				
19.Générale	fixe	mobile		
20.Mobilité	lente	modérée	rapide	
21.Angle de prise de vue (plan horizontal)	face	¾ avant - ¾ arrière	profil	dos
22.Angle de prise	contre-	plongée (vue aérienne)		

de vue (plan vertical)	plongée				
23.Cadrage	gros plan	plan taille	plan américain (coupe au niveau du colt) - plan moyen-plan pied (personnage aussi important que le décor) - plan italien (coupe dessous genou)	plan d'ensemble-plan large (personnage et décor dans le cadre sans rien autour)	plan général (personnage noyé dans le décor)
24.Nb de cuts	faible		modéré	fort	
25.Enchainements cuts	faible		modéré	fort	
26. Zooms	faible		modérée	forte	
27.Rotation caméra	oui		non	-	
28.Profondeur de champ (netteté)	flou		courte -grande	travelling	

ANNEXE 7-B

Caractérisation des séquences de test - Expérimentation B1

Distribution des couples de séquences, issus de chaque contenu, représentatives des cinq descripteurs sémantiques. Les niveaux sont indiqués comme suit: F pour Fort, M pour Modéré, Fa pour Faible puis A pour Audio, V pour Vidéo et AV pour AudioVidéo. HC représente le mode *hors-champ* du descripteur *Relation AV*. La numérotation des séquences correspond à leur ordre d''apparition au sein du contenu d'origine.

Descripteurs Sémantiques		Descripteurs Sémantiques	Relation AV				Expression Sonore			Nombre Personnages				Dynamique				Modalité		
	Contenu		Son In	Son Off	Son In	Son HC	Parole	Musique	Parole	Bruit	Fort	Modéré	Faible	Forte	Modérée	Faible	Forte	V	AV	A
			Doc._3	Doc._4	Opé_3	Opé_1	Doc._2	Doc._1	Théâ_1	Théâ_2	Danse_3	Danse_4	Opé_4	Sport_4	Sport_3	Danse_2	Danse_1	Théâ_3	Théâ_4	Sport_1	Sport_2
Luminosité	Faible		X	X			X	X	X	X	X	X				X	X	X	X	X	X
	Modérée				X	X							X								
	Forte													X	X					X	X
Couleur	Chaude										X	X				X					
	Jour		X	X	X	X	X	X					X	X	X					X	X
	Froide								X	X								X	X		
Dyn.cam	Faible		X	X	X	X	X	X	X	X	X	X	X			X	X	X	X	X	X
	Modérée				X	X															
	Forte													X	X					X	X
Détail	Faible																				
	Modéré		X	X	X	X	X	X	X	X	X	X	X			X	X	X	X	X	X
	Fort													X	X					X	X

ANNEXE 7-C

Questionnaire d'évaluation - Expérimentation B1

1) Comment évaluez-vous la qualité **audiovisuelle globale** ? (cerclez un numéro)

9 Excellent
8
7 Bon
6
5 Satisfaisant
4
3 Médiocre
2
1 Mauvais

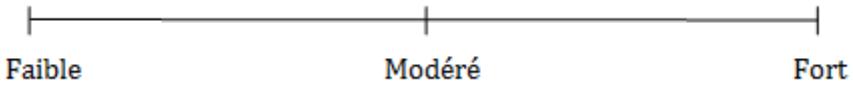
2) Comment évaluez-vous la qualité **vidéo** de l'extrait? (cerclez un numéro)

9 Excellent
8
7 Bon
6
5 Satisfaisant
4
3 Médiocre
2
1 Mauvais

3) Comment évaluez-vous la qualité **audio** de l'extrait? (cerclez un numéro)

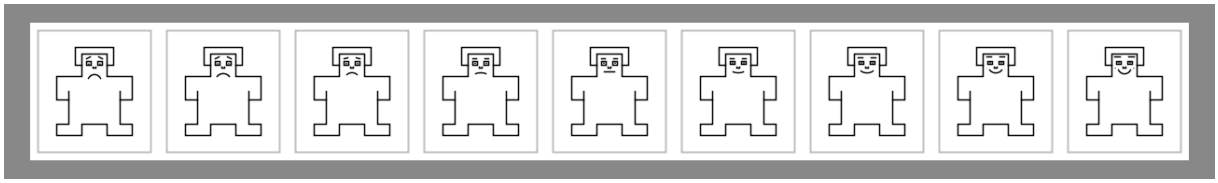
9 Excellent
8
7 Bon
6
5 Satisfaisant
4
3 Médiocre
2
1 Mauvais

4) Votre intérêt vis-à-vis de cet extrait :

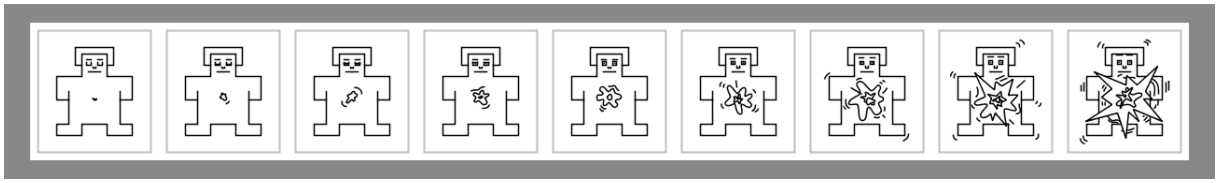


5) Pouvez-vous évaluer les émotions que vous avez ressenties lors de la visualisation et l'écoute de cet extrait, en cochant pour chaque échelle le graphique le plus représentatif :

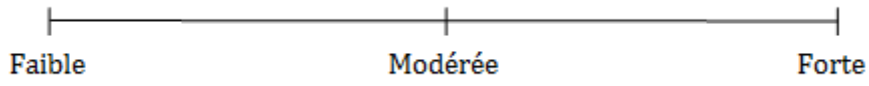
Echelle Déplaisir/Plaisir



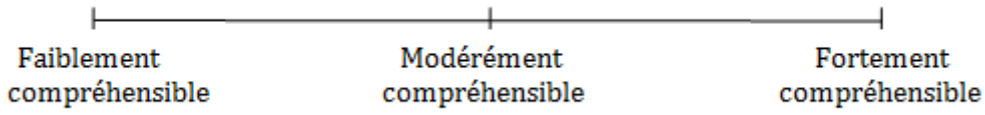
Echelle Calme/Excité



6) La quantité d'information de cet extrait était-elle:



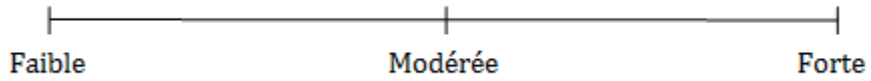
7) Ce contenu, vous a-t-il paru :



8) Selon vous, quelle est la modalité dominante de cet extrait ? (modalité qui véhicule l'information primordiale, l'absence de l'autre modalité ne gênant pas la compréhension)

- ☐ Audio
- ☐ Vidéo
- ☐ Sinon AV (dans le cas où selon vous, les deux modalités sont indispensables à la compréhension)

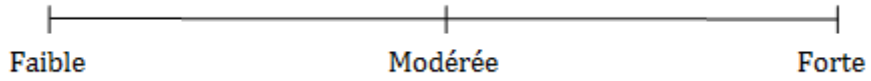
9) Selon vous, la dynamique de ce contenu était-elle :



10) Selon vous, la couleur dominante de cet extrait était :

- ☐ Chaude
- ☐ Jour
- ☐ Froide

11) Selon vous, la luminosité était-elle :



12) Exprimer votre ressenti quant à l'extrait visualisé

ANNEXE 7-D

Caractérisation naïve des séquences - Expérimentation B1

Tableau présentant la caractérisation suite à l'évaluation du panel de participant (naïfs) selon le calcul du mode réalisé pour chaque séquences et chaque descripteurs à savoir **Qualité** (MOSAV, MOSV, MOSA), **Hédonique** : Intérêt (Int), Valence (Val) et Arousal (Ar) ; **Sémantique** : Quantité d'information (Info), Compréhension (Comp), Modalité (MOD) et Dynamique (Dyn) et **Technique** : Couleur (Col) et Luminosité (Lum) ; et selon les niveaux Faible (Fa), Modéré (M), Fort (F) ou Audio (A), Vidéo (V), AudioVisuel (AV) ou Chaude (C), Jour (J), Froide.

Séquences	MOSAV	MOSV	MOSA	Int	Val	Ar	Info	Comp	Mod	Dyn	Col	Lum
Danse-1	5,43	5,64	5,57	M	3	5	Fa	Fa	V	M	C	Fa
Danse-2	6,61	6,39	6,86	F	7	7	M	F	AV	F	C	Fa
Danse-3	6,93	6,86	6,71	F	7	6	M	F	V	F	C	M
Danse-4	6,96	6,93	6,68	F	7	7	M	F	AV	F	C	Fa
Opéra-1	5,93	5,71	5,93	Fa	3	2	Fa	Fa	A	Fa	C	F
Opéra-2	6,00	6,04	6,50	Fa	4	4	M	M	A	Fa	C	M
Opéra-3	5,96	6,11	5,82	Fa	3	3	Fa	M	A	Fa	C	F
Opéra-4	6,18	6,43	6,64	Fa	4	3	Fa	Fa	A	Fa	J	F
Théâtre-1	6,79	6,75	6,25	M	6	5	M	M	A	M	C	Fa
Théâtre-2	6,64	6,79	6,14	M	5	5	M	F	V	M	C	Fa
Théâtre-3	6,57	6,57	6,50	M	5	3	Fa	M	V	Fa	C	M
Théâtre-4	6,89	6,89	6,82	M	5	5	M	M	V	M	C	M
Doc-1	6,75	7,39	6,29	M	3	5	Fa	Fa	V	M	F	M
Doc-2	6,32	6,54	6,29	M	5	3	M	F	A	M	F	M
Doc-3	6,68	6,50	6,36	M	5	5	M	F	A	Fa	F	M
Doc-4	5,89	5,79	6,39	M	5	5	M	F	A	Fa	F	Fa
Sport-1	6,39	6,46	6,46	F	7	5	F	F	V	M	J	F
Sport-2	6,82	6,93	6,18	F	7	6	M	F	V	M	J	F
Sport-3	6,64	6,64	6,61	F	6	5	F	F	V	M	J	F
Sport-4	5,54	4,93	5,93	F	7	7	M	F	V	F	J	M

ANNEXE 7-E

Tableaux de résultats (ANOVAs) -Expérimentation B1

Effet de la variable «Séquence» et de la variable aléatoire «Participant », sur chacun des descripteurs évalués par le panel de participants. Le second tableau présente les résultats du test du Khi-deux de Pearson pour les variables catégorielles « Modalité » et « Couleur ».

Variable Indépendante	Variables dépendantes	Somme des carrés	ddl effet	ddl erreur	Moyenne des carrés	F	*p*
Séquences	**Qav**	115,71	19	513	6,09	4,42	< 0, 001
	Qv	173,31	19	513	9,12	5,76	< 0, 001
	Qa	63,36	19	513	3,33	1,83	< 0,05
	Intérêt	58,18	19	513	3 ,06	6,69	< 0, 001
	Quant. Info	56,44	19	513	2,97	7,72	< 0, 001
	Compréhension	123,96	19	513	6,52	19,24	< 0, 001
	Dynamique	147,14	19	513	7,74	36,83	< 0, 001
	Luminosité	131,47	19	513	6,92	25,64	< 0, 001
	Valence	531,95	19	513	28	8,98	< 0, 001
	Arousal	456,31	19	513	24,02	8,38	< 0, 001
Participant	**Qav**	1142,09	27	513	42,30	30,73	< 0, 001
	Qv	1278,49	27	513	47,35	29,90	< 0, 001
	Qa	994,49	27	513	36,83	20,17	< 0, 001
	Intérêt	33,70	27	513	1,25	2,73	< 0, 001
	Quant. Info	43,32	27	513	1,72	4,46	< 0, 001
	Compréhension	43,70	27	513	1,62	4,77	< 0, 001
	Dynamique	17,85	27	513	0,66	3,14	< 0, 001
	Luminosité	46,64	27	513	1,73	6,40	< 0, 001
	Valence	445,23	27	513	16,49	5,29	< 0, 001
	Arousal	443,46	27	513	16,42	5,73	< 0, 001

Variable testée	Variables testées	Test	ddl	Valeur	*p*
Séquences	**Modalité**	χ^2	38	331,82	< 0,001
	Couleur	χ^2	38	326, 57	< 0,05

ANNEXE 7-F

Tableaux de contingence Modalité*Dynamique
Expérimentation B1

Tableaux de contingence relatifs à la réalisation du Khi-deux de Pearson pour le test de l'hypothèse d'indépendance entre les descripteurs *Modalité* et *Dynamique*.

Descripteurs	Modes	Dynamique			Total
		Faible	Modérée	Forte	
Modalité	AV	0	0	2	2
	V	1	7	2	10
	A	6	2	0	8
	Total	7	9	4	20

ANNEXE 7-G

Conditions de visualisation et d'écoute - Expérimentation B2

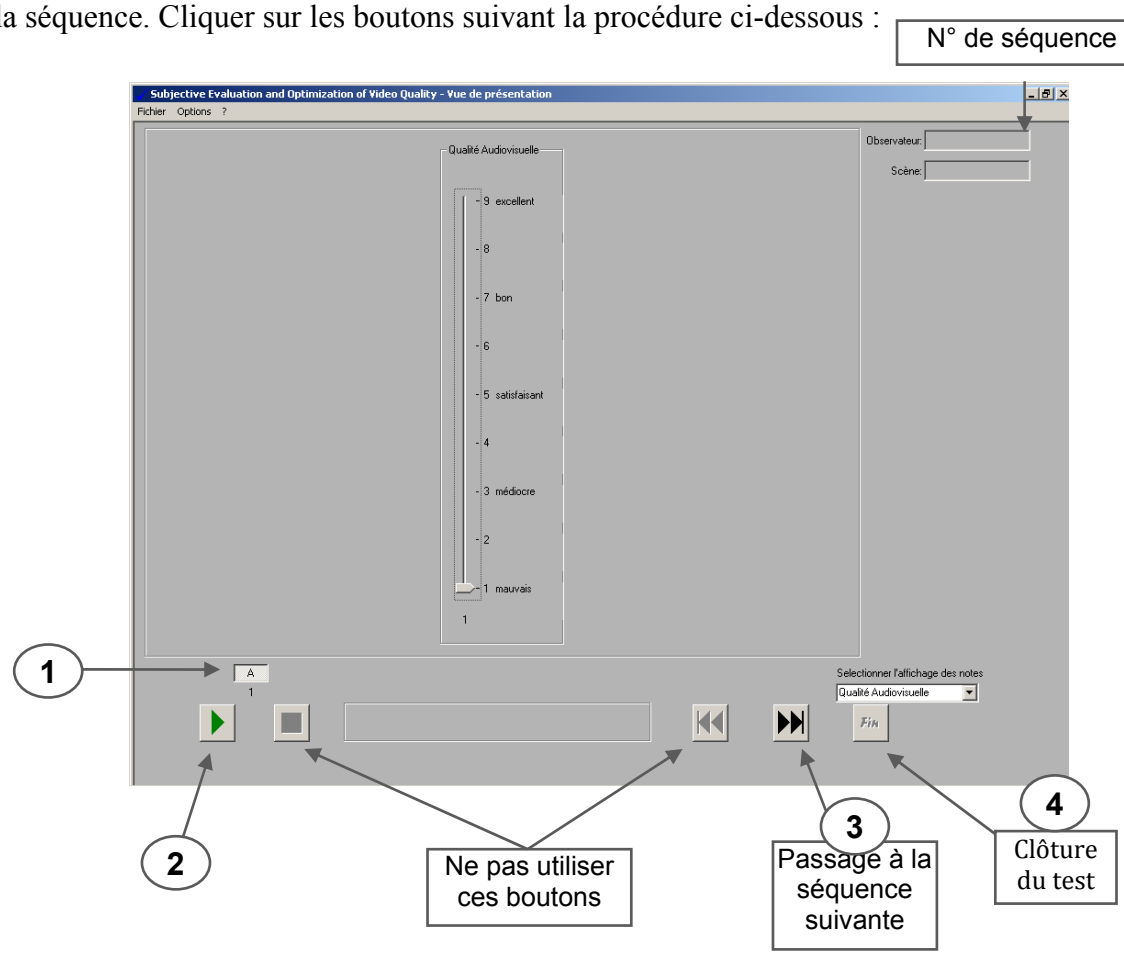
Correspondances entre les conditions de visualisation et d'écoute recommandées par la norme UIT-T P.911 et celles de l'expérimentation B2. Les dimension de la pièce sont exprimées de la manière suivante (longueur)×(Largeur)×(Hauteur).

PARAMETRES	REGLAGES P.911	REGLAGES EXPE.A
Dimensions de la pièce	l×L×H	250×310×320 cm
Distance de visualisation	De 1 à 8 H	Respecté (3,2 H)
Luminance de l'écran (valeur de crête)	de 100 à 200 cd/m2	Respecté
Rapport de luminance d'écran inactif à luminance de crête	≤ 0,05	Respecté
Rapport de luminance de l'écran au niveau de crête du blanc (lors de l'affichage d'un niveau de noir total dans une salle complètement obscure)	≤ 0,1	Respecté
Rapport entre luminance de l'arrière-fond du moniteur d'image à la valeur de crête de la luminance d'image	≤ 0,2	Respecté
Chromaticité de l'arrière-fond	D65	Respecté
Eclairement lumineux d'ambiance de la salle	≤ 20 lx	Respecté
Niveau de bruit de fond	≤ 30 dBA	Respecté
Niveau d'écoute	~ 80 dBA	Respecté
Durée de réverbération	< 500 ms, $\forall f > 150$ Hz	Respecté

ANNEXE 7-H

Consignes – Expérimentation B2

Bienvenue à FT R&D, Orange Labs. Vous allez participer à une évaluation concernant la qualité de séquences Audiovisuelles (*200).
Chaque séquence est diffusée pendant 10 secondes. Observez et écoutez avec attention toute la séquence. Cliquer sur les boutons suivant la procédure ci-dessous :



Vous devrez ensuite juger (~5 Sec.) la qualité globale, **audio et vidéo** combinée, de la séquence, en plaçant le curseur de l'échelle de Qualité Audiovisuelle, sur l'un des **neuf** niveaux ci-dessous :

9 Excellent
8
7 Bon
6
5 Satisfaisant
4
3 Médiocre
2
1 Mauvais

<u>Vous ne devez visionner chaque séquence qu'une seule fois.</u>

Passer ensuite à la séquence suivante en cliquant sur le  bouton

Lorsque vous aurez jugé l'ensemble des séquences,  cliquez sur pour terminer le test

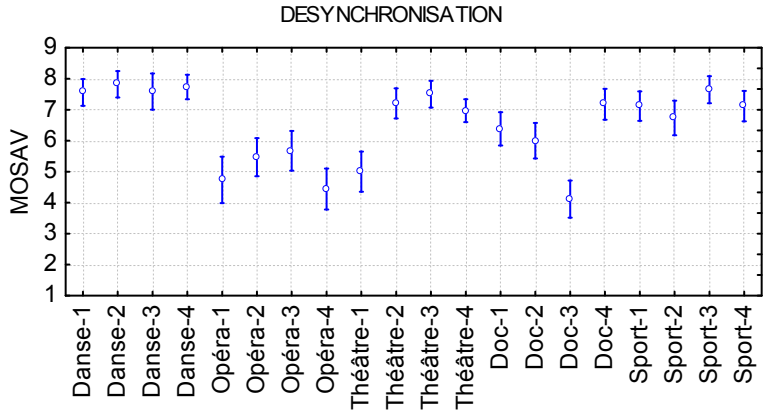
Il est indispensable de rester attentif en permanence pendant la totalité des périodes de visualisation / écoute.

Se placer en face de la marque jaune
Vous pouvez faire une pause si besoin
Nous vous remercions de votre collaboration.

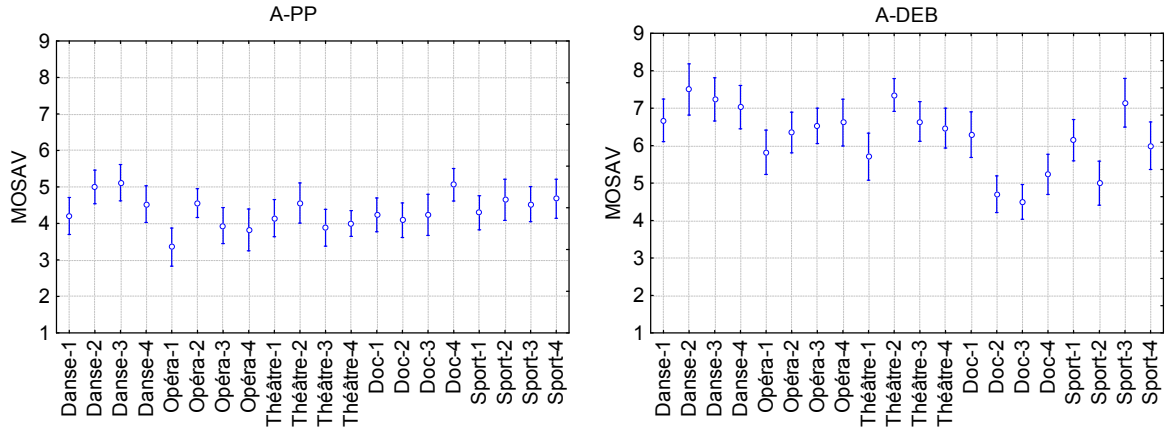
ANNEXE 7-I

Effet de type de dégradation - Expérimentation B2

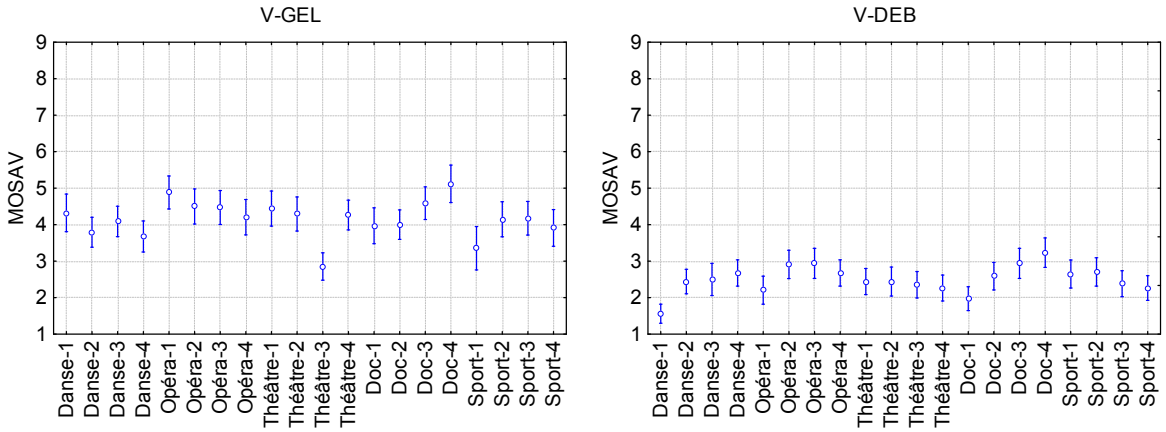
Les Figures ci-dessous illustrent les notes MOSAV, obtenues pour chaque séquence (× 20) extraite des contenus Danse, Opéra, Théâtre, Documentaire et Sport, présentées pour chaque type de dégradation (Désynchronisation, A-PP, A-DEB, V-GEL, V-DEB, V-DEB*A-PP, V-DEB*A-DEB, V-GEL*A-PP, V-GEL*A-DEB). Cette présentation permet de mieux observer certains effets discutés lors de l'analyse des résultats.



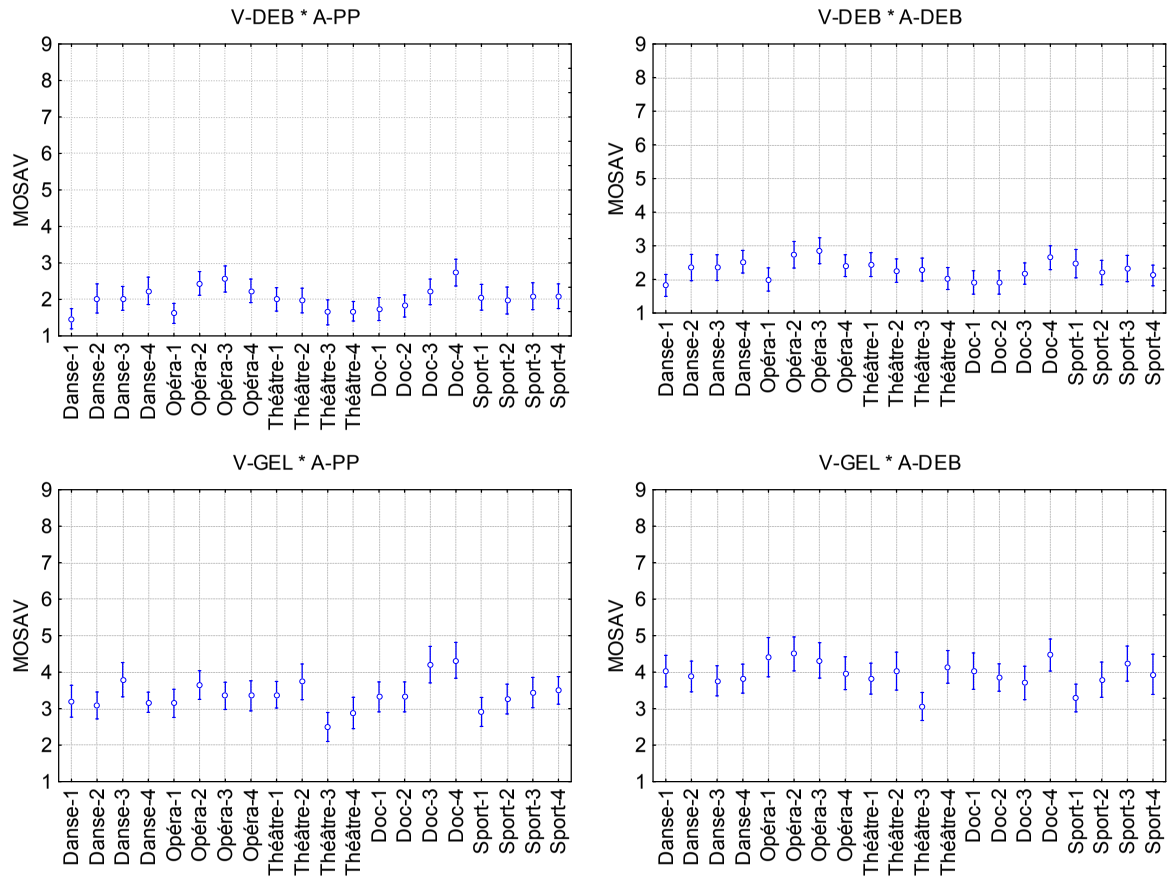
MOSAV obtenues pour chaque séquence présentée avec la dégradation *Désynchronisation*.



MOSAV obtenues pour chaque séquence présentée avec la dégradation A-PP et A-DEBIT.



MOSAV obtenues pour chaque séquence présentée avec la dégradation V-GEL et V-DEBIT.



MOSAV obtenues pour chaque séquence présentée avec les dégradations AV : V-DEB*A-PP, V-DEB*A-DEB, V-GEL*A-PP, V-GEL*A-DEB.

ANNEXE 8-A

Conditions de visualisation et d'écoute - Expérimentation C

Correspondances entre les conditions de visualisation et d'écoute recommandées par la norme UIT-T P.911 et celles de l'expérimentation C. Les dimension de la pièce sont exprimées de la manière suivante : (longueur)×(Largeur)×(Hauteur).

PARAMETRES	REGLAGES P.911	REGLAGES EXPE.A
Dimensions de la pièce	l×L×H	193×376×505 cm
Distance de visualisation	De 1 à 8 H	Respecté (4,5 H)
Luminance de l'écran (valeur de crête)	de 100 à 200 cd/m2	Respecté
Rapport de luminance d'écran inactif à luminance de crête	≤ 0,05	Respecté
Rapport de luminance de l'écran au niveau de crête du blanc (lors de l'affichage d'un niveau de noir total dans une salle complètement obscure)	≤ 0,1	Respecté
Rapport entre luminance de l'arrière-fond du moniteur d'image à la valeur de crête de la luminance d'image	≤ 0,2	Respecté
Chromaticité de l'arrière-fond	D65	Respecté
Eclairement lumineux d'ambiance de la salle	≤ 20 lx	Respecté
Niveau de bruit de fond	≤ 30 dBA	Respecté
Niveau d'écoute	~ 80 dBA	Respecté
Durée de réverbération	< 500 ms, ∀f > 150 Hz	Respecté

ANNEXE 8-B

Questionnaire - Expérimentation C

1) Comment évaluez-vous la qualité **audiovisuelle globale** ? (cerclez un numéro)

9 Excellent
8
7 Bon
6
5 Satisfaisant
4
3 Médiocre
2
1 Mauvais

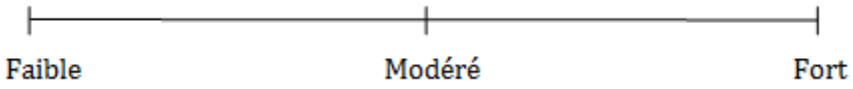
2) Comment évaluez-vous la qualité **vidéo** de l'extrait? (cerclez un numéro)

9 Excellent
8
7 Bon
6
5 Satisfaisant
4
3 Médiocre
2
1 Mauvais

3) Comment évaluez-vous la qualité **audio** de l'extrait? (cerclez un numéro)

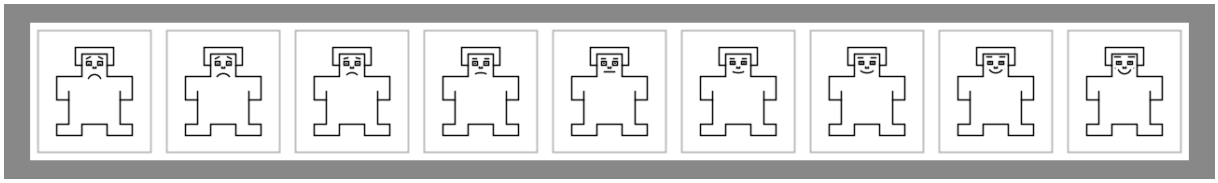
9 Excellent
8
7 Bon
6
5 Satisfaisant
4
3 Médiocre
2
1 Mauvais

4) Cochez la case correspondant à votre intérêt vis-à-vis de cet extrait :

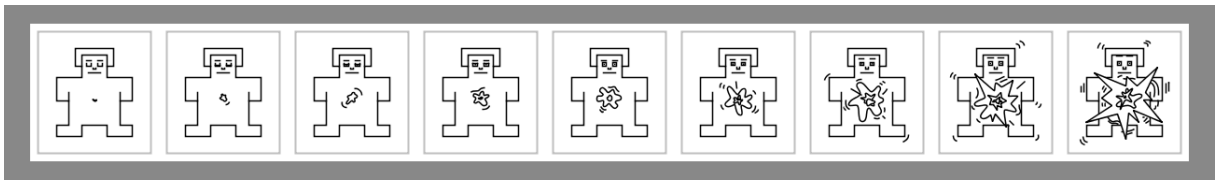


5) Pouvez-vous évaluer les émotions que vous avez ressenties lors de la visualisation et l'écoute de cet extrait, en cochant pour chaque échelle le graphique le plus représentatif :

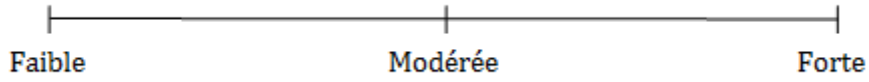
Echelle Déplaisir/Plaisir



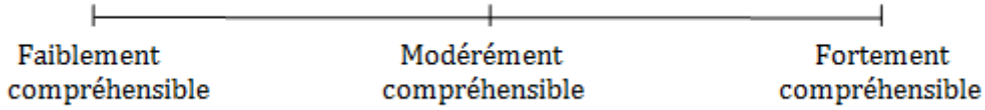
Echelle Calme/Excité



6) La quantité d'information de ce contenu était-elle:



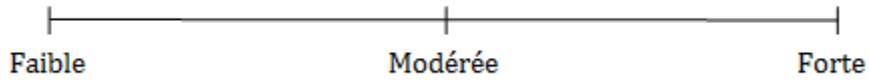
7) Ce contenu vous a-t-il paru :



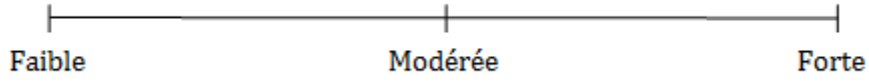
8) Selon vous, quelle est la modalité dominante de cet extrait ? (modalité qui véhicule l'information primordiale, l'absence de l'autre modalité ne gênant pas la compréhension)

- ☐ Audio
- ☐ Vidéo
- ☐ Sinon AV (dans le cas où, selon vous, les deux modalités sont indispensables à la compréhension)

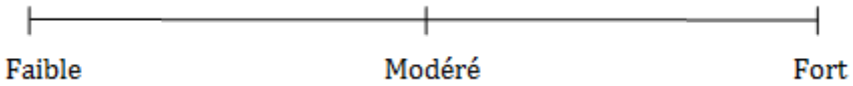
9) Selon vous, la dynamique de ce contenu était-elle :



10) Selon vous, la luminosité était-elle :



11) Avez-vous perçu des dégradations sur le son seul : ☐ Oui ☐ Non
Quel est le degré de certitude associé à votre réponse ?



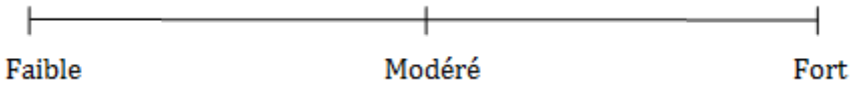
Si (et seulement si) vous avez perçu des dégradations sur le son seul :
Ces dégradations ont-elles eu un impact sur la compréhension du contenu ?

- ☐ Pas du tout
- ☐ Légèrement
- ☐ Moyennement
- ☐ Beaucoup
- ☐ Extrêmement

Ces dégradations ont-elles suscité chez vous des émotions négatives telles que de l'agacement, de l'énervement, du stress, de la frustration, … ?

- ☐ Pas du tout
- ☐ Légèrement
- ☐ Moyennement
- ☐ Beaucoup
- ☐ Extrêmement

12) Avez-vous perçu des dégradations sur <u>l'image seule</u> : ☐ Oui ☐ Non
Quel est le degré de certitude associé à votre réponse ?



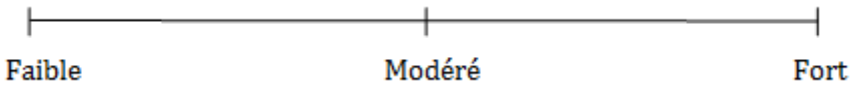
Si (et seulement si) vous avez perçu des dégradations sur l'image seule :
Ces dégradations ont-elles eu un impact sur la compréhension du contenu ?

- ☐ Pas du tout
- ☐ Légèrement
- ☐ Moyennement
- ☐ Beaucoup
- ☐ Extrêmement

Ces dégradations ont-elles suscité chez vous des émotions négatives telles que de l'agacement, de l'énervement, du stress, de la frustration, … ?

- ☐ Pas du tout
- ☐ Légèrement
- ☐ Moyennement
- ☐ Beaucoup
- ☐ Extrêmement

Avez-vous perçu des dégradations sur <u>l'image et le son en même temps</u> : ☐ Oui ☐ Non
Quel est le degré de certitude associé à votre réponse ?



Si (et seulement si) vous avez perçu des dégradations sur l'image et le son en même temps :
Ces dégradations ont-elles eu un impact sur la compréhension du contenu ?

- ☐ Pas du tout
- ☐ Légèrement
- ☐ Moyennement
- ☐ Beaucoup
- ☐ Extrêmement

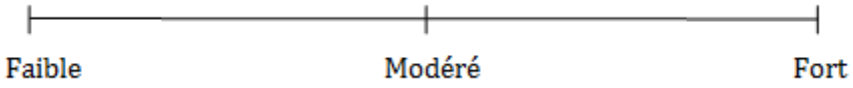
Ces dégradations ont-elles suscité chez vous des émotions négatives telles qu'agacement, énervement, stress, frustration, …

- ☐ Pas du tout
- ☐ Légèrement
- ☐ Moyennement
- ☐ Beaucoup
- ☐ Extrêmement

☐ oui ☐ non

13) Avez-vous perçu un décalage temporel entre l'image et le son :

Quel est le degré de certitude associé à votre réponse ?



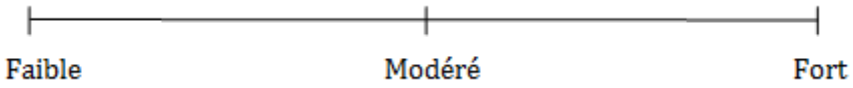
Si (et seulement si) vous avez perçu un décalage temporel entre l'image et le son :
Ce décalage temporel image/son a-t-il eu un impact sur la compréhension du contenu ?

- ☐ Pas du tout
- ☐ Légèrement
- ☐ Moyennement
- ☐ Beaucoup
- ☐ Extrêmement

Ce décalage temporel image/son a-t-il suscité chez vous des émotions négatives telles que de l'agacement, de l'énervement, du stress, de la frustration, … ?

- ☐ Pas du tout
- ☐ Légèrement
- ☐ Moyennement
- ☐ Beaucoup
- ☐ Extrêmement

14) Avez-vous ressenti une gêne liée à la 3D : ☐ oui ☐ non
Quel est le degré de certitude associé à votre réponse ?



Si (et seulement si) vous avez été gêné(e) par la 3D:
Cette gêne a-t-elle eu un impact sur la compréhension du contenu ?

- ☐ Pas du tout
- ☐ Légèrement
- ☐ Moyennement
- ☐ Beaucoup
- ☐ Extrêmement

Cette gêne a-t-elle suscité chez vous des émotions négatives telles que de l'agacement, de l'énervement, du stress, de la frustration, … ?

- ☐ Pas du tout
- ☐ Légèrement

- Moyennement
- Beaucoup
- Extrêmement

15) Exprimez votre ressenti quant à l'extrait visualisé/entendu

---

ANNEXE 8-C

Evaluation du niveau de fatigue - Expérimentation C

Echelle de fatigue à 7 niveaux sélectionnée sur la base des travaux de McAuley et Courneya (1994 [48]) autour de la mesure de réponses psychologiques consécutives à un exercice physique.

En entourant un numéro de l'échelle ci-dessous, veuillez indiquer votre degré de fatigue :

1	2	3	4	5	6	7
Pas du tout			Modérément			Extrêmement

[48] McAuley, E. et Courneya, K. S. (1994). The subjective exercise experiences scale (SEES): Development and preliminary validation. *Journal of Sport and Exercise Psychology, 16*, 163-177.

ANNEXE 8-D

Synopsis - Expérimentation C

Synopsis pour chacun des cinq contenus de test évalués dans l'expérimentation C à savoir Danse(Balé de Rua), Opéra (Don Giovanni), Documentaire (Mormeck), Sport (finale de tennis Rolland Garros) et de Théâtre (fourberies de Scapin).

Danse : extrait de du ballet *Balé de Rua*

Révélation de la Biennale de la Danse de Lyon en 2002, le Balé de Rua (« Ballet de la Rue ») est issu des danses de rue nord-américaines, de la capoeira et de la samba. Le langage de ses interprètes se nourrit également de leurs racines africaines et de leur quotidien, la plupart d'entre eux ayant vécu de petits métiers dans les favelas.

Sur des musiques originales et certains grands airs brésiliens, les danseurs du Balé de Rua s'accompagnent eux-mêmes à grand renfort de percussions. Doués d'une prodigieuse ingéniosité, et d'une énergie communicative, ils délivrent leur message d'espoir et de joie à travers les chorégraphies de Marco Antônio Garcia.

Le Balé de Rua raconte une histoire afro-brésilienne, celle d'un groupe issu des quartiers populaires d'une petite ville brésilienne, celle d'amis qui repeignent le monde tout en couleurs, grâce à la magie du rêve et de la danse, un monde à l'image du Brésil.

Opéra : extrait de la pièce *Don Giovanni*

Don Giovanni (issu de la pièce *Don Juan* de Molière), est un opéra de Mozart en deux actes et en langue italienne. Cet opéra est aujourd'hui considéré comme un des opéras majeurs de Mozart.

L'extrait qui vous sera présenté montre l'acte 1 où Don Giovanni, en fuite pour avoir déshonoré Donna Elvira en l'épousant puis en la délaissant au profit d'autres conquêtes, est sauvé de la noyade par Masetto, un paysan. Don Giovanni fait alors la rencontre de Zerlina, qui lui plaît, et se débarrasse du fiancé jaloux, qui n'est autre que Masetto. Dès que Don Giovanni se retrouve seul avec Zerlina, il commence à la séduire. A la fin de l'acte, Donna Elvira les rejoint et emmène Zerlina juste avant qu'elle ne cède aux avances de Don Giovanni.

Théâtre : extrait de la pièce *Les Fourberies de Scapin*

Les Fourberies de Scapin est une comédie de Molière en trois actes. En l'absence de leurs parents partis en voyage, Octave, fils d'Argante, et Léandre, fils de Géronte, se sont respectivement épris de Hyacinte (jeune fille pauvre et de naissance inconnue qu'Octave vient secrètement d'épouser) et de Zerbinette (une jeune esclave égyptienne).

Octave et Léandre appréhendent le retour de leurs parents car ces derniers envisageaient pour eux d'autres projets de mariages. Entretemps, Zerbinette se fait enlever par les Égyptiens qui demandent alors une rançon à Léandre. Afin de se sortir de ce mauvais pas, tous deux font appels à Scapin, le serviteur de Léandre, réputé pour sa ruse. Mais Scapin indiscret laisse échapper le secret et Géronte la mère de Léandre, apprend l'engagement de son fils pour Zerbinette. Très en colère, Géronte accueille très mal cette nouvelle.

Pour punir Scapin de son indiscrétion, Léandre le corrige sévèrement. Scapin décide alors de se venger de Géronte. Il persuade la vieille femme de se cacher dans un sac et la roue de coups. Géronte s'en sort avec une jambe cassée.

Puis, à force de mensonges et de fourberies, Scapin réussi à extorquer à Argante, père d'Octave, la somme nécessaire pour payer la rançon de Zerbinette. SScapin paierait cher ses ruses si une double reconnaissance ne révélait en Hyacinte la fille perdue de Géronte et en Zerbinette celle d'Argante.
L'extrait qui vous sera présenté montre l'acte 3 où les fourberies de Scapin sont révélées à Argante et Géronte.

Documentaire : documentaire entier sur le boxeur *Jean-Marc Mormeck*

Documentaire de douze minutes sur le boxeur Français Jean-Marc Mormeck qui a remporté à deux reprises le titre de champion du monde de boxe en poids lourds et légers.

Sport : extrait de la finale de tennis *Rolland Garros 2011*

Extrait du 4ème jeu de la finale Hommes de Rolland Garros 2011 opposant Roger Federer et Rafael Nadal. Dans l'extrait qui sera présenté, Rafael Nadal mène 15-0.

ANNEXE 8-E

Evaluation de préférence - Expérimentation C

Echelle permettant le classement des cinq contenus, après lecture des synopsis, selon la préférence du participant. Cette évaluation était également proposée après la visualisation des contenus.

Veuillez classer de 1 à 5, selon votre préférence, les extraits dont le résumé vous a été présentés. Merci de reporter votre classement dans les cases ci-dessous, 1 étant l'extrait préféré et 5 le moins préféré :

☐ Balé de Rua

☐ Don Giovanni

☐ Fourberies de Scapin

☐ Mormeck

☐ Roland Garros

ANNEXE 8-F

Consignes - Expérimentation C

Bonjour et merci de votre participation.

Dans le cadre de l'évaluation d'un nouveau service de diffusion à l'essai de TV3D de contenus AV, vous allez visualiser **5 extraits audiovisuels** différents :

- Opéra (extrait de Don Giovanni)
- Théâtre (extrait des Fourberies de Scapin)
- Danse (extrait du Balé de rua)
- Sport (extrait de la finale de Rolland Garros 2011)
- Documentaire (sur le boxer français Jean-Marc Mormeck)

Ces 5 extraits vous seront présentés successivement. Entre chaque extrait vous disposerez d'une **pause de 5 minutes** (signalée par un écran noir) pour compléter un **questionnaire** (3 pages recto verso) à propos de l'extrait que vous aurez vu.

Au total, vous devrez donc remplir 5 questionnaires, un pour chaque extrait.

Le début de l'extrait suivant sera toujours signalé une minute (affichage du message « *début du contenu dans une minute* ») et six secondes avant (décompte 5, 4, 3, 2, 1, 0) vous avertissant ainsi de la reprise de la visualisation.

Questionnaire :

- Pour chaque extrait, il vous sera demandé d'en juger la qualité en entourant, pour chacune des trois échelles présentées (qualité audiovisuelle, qualité vidéo, qualité audio), l'un des neuf niveaux reflétant votre opinion.

- Le questionnaire présentera également une série d'échelles afin de recueillir votre ressenti et votre perception quant à l'extrait visualisé.

Merci de prendre connaissance du questionnaire après la lecture des consignes.

Observez et écoutez avec attention l'ensemble de l'extrait avant d'exprimer votre jugement.

<u>Consignes de posture</u> :
Il vous sera demandé durant le test de :

- **maintenir** autant que possible la **posture confortable** que vous aurez choisie en début de test
- **ne pas positionner votre main libre devant votre bouche**
- garder la **main** choisie pour l'installation des capteurs, **la plus immobile** possible

ANNEXE 8-G

Tableaux de répartition des effectifs pour le classement de préférences -Expérimentation C

Tableau de répartition des effectifs pour chaque position du classement des contenus par ordre de préférence avant et après le test.

Position / Contenu	1		2		3		4		5	
	avant	**après**	**avant**	**après**	**avant**	**après**	**avant**	**après**	**avant**	**après**
Danse	11	15	6	10	9	6	3	2	4	1
Doc.	7	7	6	6	5	6	6	11	9	4
Opéra	3	2	9	2	6	4	8	2	7	24
Sport	4	5	7	12	5	12	9	3	8	2
Théâtre	8	5	5	4	8	6	7	16	5	3

ANNEXE 8-H

Tableaux de pourcentage de détection et de niveau de certitude Expérimentation C

Tableau de pourcentage de détection de la dégradation (oui/non) et de chaque niveau de certitude Faible (Fa), Modéré (M) ou Fort (F) associé pour chaque type de dégradation (audio, vidéo, audio-vidéo, désynchronisation -Désynchro.- et Gêne 3D) et chaque contenu (Opéra, Danse, Sport, Documentaire -Doc.- et Théâtre).

		oui/non		**Certitude *oui***			**Certitude *non***		
Dégradation	**Contenu**	**oui**	**non**	**Fa**	**M**	**F**	**Fa**	**M**	**F**
Audio	**Opéra**	87,9	12,1	13,8	17,2	69,0	25	50	0
	Danse	96,9	3,1	16,1	25,8	58,1	0	0	100
	Sport	94,1	5,9	6,3	15,6	78,1	0	50	0
	Doc.	93,9	6,1	0	28,1	71,9	0	0	0
	Théâtre	93,9	6,1	3,2	16,1	80,6	0	50	0
	Total	93,9	6,7	7,7	20,6	71,6	10	40	10
Vidéo	**Opéra**	93,9	6,1	3,3	3,3	93,3	0	0	100
	Danse	100	0	0	21,9	78,1	0	0	0
	Sport	100	0	0	17,6	82,4	0	0	0
	Doc.	100	0	0	6,1	93,9	0	0	0
	Théâtre	100	0	3,0	9,1	87,9	0	0	0
	Total	98,8	1,2	1,2	11,7	87,0	0	0	0
AudioVidéo	**Opéra**	75	25	21,7	21,7	56,5	42,9	14,3	0
	Danse	72,7	27,3	25	20,8	54,2	60	10	0
	Sport	76,5	23,5	19,2	30,8	50	37,5	25	12,5
	Doc.	80,6	19,4	20	16	64	66,7	0	33,3
	Théâtre	78,8	21,2	19,2	26,9	53,8	20	40	0
	Total	76,7	23,3	21,0	23,4	55,6	47,2	16,7	8,3
Désynchro.	**Opéra**	91,2	8,8	10	20	70	0	50	50
	Danse	5,9	94,1	0	50	50	15,6	12,5	31,3
	Sport	2,9	97,1	0	100	0	24,2	18,2	18,2
	Doc.	88,2	11,8	0	10	90	25	25	25
	Théâtre	78,8	21,2	7,7	15,4	76,9	28,6	14,3	0
	Total	53,3	46,7	5,6	16,9	77,5	20,5	16,7	23,1
Gêne 3D	**Opéra**	51,5	48,5	18,8	50	31,3	0	6,3	37,5
	Danse	24,2	75,8	50	37,5	12,5	4	12	40
	Sport	44,1	55,9	20	40	40,0	0	26,3	36,8
	Doc.	23,5	76,5	37,5	12,5	50,0	11,5	19,2	30,8
	Théâtre	87,9	12,1	13,8	24,1	62,1	0	75	25
	Total	46,1	53,9	22,4	32,9	44,7	4,4	18,9	35,6

ANNEXE 8-I

Tableaux de résultats (ANOVAs) - Expérimentation C

Tableau représentant les effets principaux des variables indépendantes (VI) « Contenu » et « Participant » (aléatoire) sur les variables dépendantes (VD) « Compréhension » (Compr.) et « Emotions négatives » (Emo -) pour chaque type de dégradation : Audio (A), Vidéo (V), AudioVidéo combinée (AV), Désynchronisation (D) et Gêne 3D (3D).

VI	VD	Somme des carrés	ddl effet	ddl erreur	Moyenne des carrés	F	*p*
Contenu	**Compr. A**	55,40	4	116	13,85	17,60	<0,001
	Emo - A	23,44	4	116	5,86	7,62	<0,001
	Compr .V	7,41	4	122	1,85	1,65	0,17
	Emo - V	2,93	4	122	0,73	0,72	0,58
	Compr .AV	14,85	4	86	3,71	3,90	<0,01
	Emo - AV	3,74	4	86	0,93	1,24	0,30
	Compr .D	6,43	4	53	1,61	2,15	0,09
	Emo - D	4,59	4	53	1,15	1,73	0,16
	Compr. 3D	3,50	4	41	0,88	1,05	0,39
	Emo - 3D	4,87	4	41	1,22	1,22	0,32
Participant	**Compr. A**	50,32	32	116	1,53	1,94	<0,01
	Emo - A	67,23	32	116	2,04	7,62	<0,001
	Compr .V	55,55	32	122	1,68	1,50	0,059
	Emo - V	69,24	32	122	2,10	2,05	<0,01
	Compr .AV	59,77	32	86	1,81	1,90	<0,01
	Emo - AV	72,99	32	86	2,21	2,93	<0,001
	Compr .D	57,39	32	53	1,79	2,39	<0,01
	Emo - D	60,50	32	53	1,89	2,84	<0,001
	Compr. 3D	29,98	32	41	0,97	1,16	0,32
	Emo - 3D	48,59	32	41	1,57	1,58	0,09

ANNEXE 8-J

Tableaux de résultats (ANOVAs) / Mesures oculaires Expérimentation C

A- Tableau représentant les effets principaux de la variable indépendante (VI) « Contenu » et de la variable aléatoire « Participant » sur les variables dépendantes (VD) « EBdur », « EBfreq », « PERCLOS », « SAC » et « DP ».

VI	VD	Somme des carrés	ddl effet	ddl erreur	Moyenne des carrés	F	*p*
Contenu	**EBdur**	0,00	4	72	0,00	2,09	0, 09
	EBfreq	0,04	4	72	0,0	0,26	0,90
	PERCLOS	0,00	4	72	0,00	0,49	0,74
	SAC	0,00	4	72	0,00	5,56	<0,001
	DP	0,00	4	72	0,00	8,13	<0,001
Participants	**EBdur**	0,04	18	72	0,00	3,07	<0,001
	EBfreq	4,38	18	72	0,24	6,82	<0,001
	PERCLOS	0,21	18	72	0,01	5,99	<0,001
	SAC	0,00	18	72	0,00	1,84	<0,05
	DP	0,00	18	72	0,00	23,76	<0,001

B- Tableau représentant les effets principaux de la variable indépendante (VI) « Période» et de la variable aléatoire « Participant » sur les variables dépendantes (VD) « EBdur », « EBfreq », « PERCLOS », « SAC » et « DP » étudiées pour chaque contenu Danse, Documentaire, Opéra, Sport et Théâtre.

VI	Contenus	VD	Somme des carrés	ddl effet	ddl erreur	Moyenne des carrés	F	*p*
Périodes	**Danse**	**EBdur**	0,09	7	126	0,01	10,63	<0,001
		EBfreq	0,41	7	126	0,06	0,72	0,66
		PERCLOS	0,02	7	126	0,00	1,28	0,27
		SAC	0,00	7	126	0,00	0,61	0,74
		DP	0,00	7	126	0,00	9,78	<0,001
	Documentaire	**EBdur**	0,05	8	144	0,00	3,59	<0,001
		EBfreq	0,47	8	144	0,06	1,19	0,31
		PERCLOS	0,00	8	144	0,00	0,93	0,50
		SAC	0,00	8	144	0,00	1,80	0,08
		DP	0,00	8	144	0,00	10,24	<0,001
	Opéra	**EBdur**	0,09	8	144	0,02	5,55	<0,001
		EBfreq	3,08	8	144	0,39	1,19	0,31
		PERCLOS	0,00	8	144	0,00	0,88	0,53

		SAC	0,00	8	144	0,00	0,80	0,60
		DP	0,00	8	144	0,00	10,58	<0,001
	Sport	**EBdur**	0,05	7	126	0,00	6,53	<0,001
		EBfreq	0,33	7	126	0,05	1,42	0,20
		PERCLOS	0,00	7	126	0,00	0,48	0,84
		SAC	0,00	7	126	0,00	1,12	0,36
		DP	0,00	7	126	0,00	16,91	<0,001
	Théâtre	**EBdur**	0,03	7	126	0,00	2,98	<0,01
		EBfreq	0,36	7	126	0,05	0,84	0,55
		PERCLOS	0,00	7	126	0,00	1,38	0,22
		SAC	0,00	7	126	0,00	3,03	<0,01
		DP	0,00	7	126	0,00	15,60	<0,001
Participant	**Danse**	**EBdur**	0,13	18	126	0,01	5,64	<0,001
		EBfreq	6,11	18	126	0,34	4,15	<0,001
		PERCLOS	1,55	18	126	0,09	42,98	<0,001
		SAC	0,01	18	126	0,00	1,83	<0,05
		DP	0,00	18	126	0,00	9,78	<0,001
	Documentaire	**EBdur**	0,35	18	144	0,02	11,82	<0,001
		EBfreq	10,07	18	144	0,56	11,36	<0,001
		PERCLOS	0,05	18	144	0,00	8,66	<0,001
		SAC	0,01	18	144	0,00	3,23	<0,001
		DP	0,00	18	144	0,00	30,76	<0,001
	Opéra	**EBdur**	0,22	18	144	0,01	5,79	<0,001
		EBfreq	12,27	18	144	0,68	2,11	<0,01
		PERCLOS	0,11	18	144	0,01	10,56	<0,001
		SAC	0,01	18	144	0,00	1,04	0,42
		DP	0,00	18	144	0,00	10,58	<0,001
	Sport	**EBdur**	0,10	18	126	0,01	4,63	<0,001
		EBfreq	11,57	18	126	0,64	19,28	<0,001
		PERCLOS	1,02	18	126	0,06	22,87	<0,001
		SAC	0,02	18	126	0,00	6,42	<0,001
		DP	0,00	18	126	0,00	12,55	<0,001
	Théâtre	**EBdur**	0,22	18	126	0,01	10,00	<0,001
		EBfreq	14,15	18	126	0,79	12,83	<0,001
		PERCLOS	0,90	18	126	0,05	53,53	<0,001
		SAC	0,00	18	126	0,00	2,00	<0,05
		DP	0,00	18	126	0,00	15,60	<0,001

ANNEXE 8-K

Tableaux de résultats (ANOVAs) / Mesures physiologiques Expérimentation C

A-Tableau représentant les effets principaux de la variable indépendante (VI) « Contenu » et de la variable aléatoire « Participant » sur les variables dépendantes (VD)) « AED*n* », « FC*n* », «TCP*n* » et « VSP*n* ».

VI	VD	Somme des carrés	ddl effet	ddl erreur	Moyenne des carrés	F	*p*
Contenu	**AED*n***	0,13	4	100	0,03	3,42	<0,05
	FC*n*	0,00	4	100	0,00	0,86	0,49
	TCP*n*	0,01	4	100	0,00	1,17	0,33
	VSP*n*	0,00	4	100	0,00	0,85	0,49
Participant	**AED*n***	3,59	25	100	0,14	15,32	<0,001
	FC*n*	0,16	25	100	0,006	7,52	<0,001
	TCP*n*	0,76	25	100	0,03	23,99	<0,001
	VSP*n*	0,00	25	100	0,00	15,79	<0,001

B- Tableau représentant les effets principaux des variables indépendantes (VI) « Période» et « Participant » (aléatoire) sur les variables dépendantes (VD) « AED », « FC», «TCP» et « VSP» étudiées pour chaque contenu de test Danse, Documentaire, Opéra, Sport et Théâtre.

VI	Contenus	VD	Somme des carrés	ddl effet	ddl erreur	Moyenne des carrés	F	*p*
Période	**Danse**	**AED*n***	1,06	7	175	0,15	23,80	<0,001
		FC*n*	0,01	7	175	0,00	1,84	0,08
		TCP*n*	0,00	7	175	0,00	1,33	0,24
		VSP*n*	0,00	7	175	0,00	0,15	0,99
	Documentaire	**AED*n***	0,51	8	200	0,06	13,75	<0,001
		FC*n*	0,03	8	200	0,00	1,55	0,14
		TCP*n*	0,00	8	200	0,00	1,14	0,34
		VSP*n*	0,00	8	200	0,00	0,27	0,97
	Opéra	**AED*n***	0,95	8	200	0,12	21,12	<0,001
		FC*n*	0,02	8	200	0,00	3,38	<0,01
		TCP*n*	0,00	8	200	0,00	2,64	<0,01
		VSP*n*	0,00	8	200	0,00	0,01	1
	Sport	**AED*n***	0,26	7	175	0,038	9,67	<0,001
		FC*n*	0,02	7	175	0,002	3,83	<0,001
		TCP*n*	0,002	7	175	0,00	1,36	0,22
		VSP*n*	0,00	7	175	0,00	1,61	0,13
	Théâtre	**AED*n***	0,21	7	175	0,03	7,89	<0,001

		FC*n*	0,01	7	175	0,00	3,68	<0,001
		TCP*n*	0,00	7	175	0,00	3,65	<0,01
		VSP*n*	0,00	7	175	0,00	1,47	0,18
Participant	**Danse**	**AED*n***	5,84	25	175	0,01	36,62	<0,001
		FC*n*	0,43	25	175	0,00	20,81	<0,001
		TCP*n*	1,68	25	175	0,00	1285,39	<0,001
		VSP*n*	0,00	25	175	0,00	0,90	0,61
	Documentaire	**AED*n***	9,68	25	200	0,00	82,96	<0,001
		FC*n*	0,37	25	200	0,00	7,05	<0,001
		TCP*n*	2,28	25	200	0,00	4887,39	<0,001
		VSP*n*	0,00	25	200	0,00	0,90	0,61
	Opéra	**AED*n***	7,49	25	200	0,01	53,25	<0,001
		FC*n*	0,42	25	200	0,00	23,82	<0,001
		TCP*n*	1,08	25	200	0,00	397,11	<0,001
		VSP*n*	0,00	25	200	0,00	1,13	0,31
	Sport	**AED*n***	8,05	25	175	0,00	82,72	<0,001
		FC*n*	0,47	25	175	0,00	33,67	<0,001
		TCP*n*	0,60	25	175	0,00	135,24	<0,001
		VSP*n*	0,00	25	175	0,00	0,65	0,90
	Théâtre	**AED*n***	6,75	25	175	0,00	71,73	<0,001
		FC*n*	0,39	25	175	0,00	31,59	<0,001
		TCP*n*	1,75	25	175	0,00	3773,59	<0,001
		VSP*n*	0,00	25	175	0,00	0,57	0,95

ANNEXE 9-A

Mesure de la perception de la qualité audiovisuelle par analyse conjointe de signaux physiologiques

JULIE LASSALLE[1], JEROME DANIEL[1], RONAN LE PAGE[2], JEAN-MARC GOUJON[2], LAETITIA GROS[1]

[1] Orange Labs
2 Avenue Pierre Marzin, 22307 Lannion Cedex, France
[2] Laboratoire Foton ENSSAT
6 rue Kérampont, BP 80518, 22305 Lannion, France

[1]julielassalle84@gmail.com, jerome.daniel@orange.com, laetitia.gros@orange.com
[2]ronan.le-page@enssat.fr, jean-marc.goujon@enssat.fr

Résumé - L'influence de la qualité audiovisuelle (AV) sur l'utilisateur a été étudiée à partir de l'analyse de mesures physiologiques complétant les mesures subjectives habituellement utilisées. Le présent papier propose une méthode d'analyse de ces signaux basée sur l'extraction d'indicateurs et la définition d'un modèle empirique de détection automatique (par fusion de données hétérogènes) des modifications éventuelles de l'activité physiologique en réaction à la présence de dégradations de qualité et plus globalement, à la présentation de stimuli AV.

Abstract - The influence of the video quality (AV) assessment was studied upon the basis of physiological measurements in addition to subjectives measurements usually used. In this paper, we propose new indicators extracted from these signals and an empirical method (with heterogenous data fusion) for automatic detection of physiological reaction due to quality degradation or more generally in response of stimuli.

1 Contexte

Dans un contexte fortement concurrentiel, un des principaux enjeux pour les acteurs du domaine de l'offre de services audiovisuels (AV) est de garantir une qualité d'expérience (QoE) optimale notamment en termes de qualité du signal audio/vidéo restitué à l'utilisateur (QAV). Actuellement, la QAV est principalement évaluée à partir de la collecte de notes données par des testeurs naïfs sur des échelles de qualité, après visualisation/écoute de séquences AV traitées par le service ou la technologie à évaluer. Ces tests subjectifs suivent des procédures recommandées par l'Union Internationale des Télécommunications (UIT). Cependant, la QoE ne peut se réduire à la seule évaluation de la qualité du signal AV restitué mais doit également tenir compte des différents facteurs capables de l'influencer tels que la fatigue ou l'effort mental induits par le niveau de qualité. Une méthode alternative pour pallier les faiblesses des méthodes subjectives actuelles pour évaluer la qualité pourrait consister à compléter les mesures subjectives par des mesures physiologiques reflétant l'activité du système nerveux autonome [1, 2, 3].

1.1 Procédure expérimentale

Trente-trois participants ont visualisé et écouté un total de cinq contenus audiovisuels (Opéra, Sport, Théâtre, Ballet et Documentaire) présentés en format 3D/stéréo. La durée des contenus de test était comprise entre 10 et 14 min. Durant la visualisation, des dégradations audio (10% de perte de paquets), vidéo (réduction du débit, variable selon le contenu) et audiovisuelles (combinaison des dégradations audio et vidéo et désynchronisation image/son avec 1500 ms d'avance du son) étaient appliquées. Chaque dégradation audio (A), vidéo, (V), audio-vidéo combinée (AV) et désynchronisation (D) était insérée pendant une minute. Chaque contenu présentait un pattern de dégradations identique pour tous les testeurs (voir Figure 1 ci-dessous).

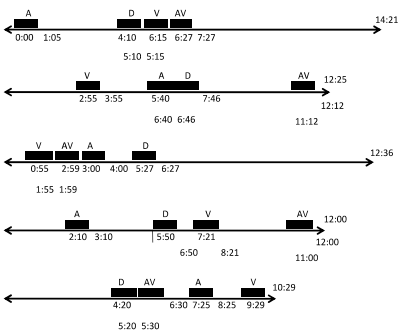


Figure 1 : signaux d'entrée du modèle, état détecté en sortie du modèle et localisation des zones de dégradations

En revanche, l'ordre de présentation des contenus était aléatoire. Après chaque visualisation, il était demandé au participant de juger les niveaux de qualité audiovisuelle, vidéo et audio sur une échelle catégorielle en neuf points et cinq items (Excellent-Bon-Satisfaisant-Médiocre-Mauvais). Au cours de l'expérience, quatre indices physiologiques ont été mesurés en continu : la fréquence cardiaque (FC), l'activité électrodermale

fréquence cardiaque (FC), l'activité électrodermale (AED, mesurée à partir de la conductance électrodermale), le volume sanguin périphérique (VSP) et la température cutanée périphérique (TCP). Ces mesures ont été recueillies à l'aide de l'outil *Procomp Infiniti* (Thought Technologies™).

1.2 Approches pour l'analyse des données

Il était attendu que la présence de dégradations A et/ou V soit reflétée par les mesures physiologiques recueillies. Une première étape a consisté à réduire les données par la moyenne de chaque signal physiologique pour chaque période : dégradée (1 min) ou non dégradée (durées variables). Une série d'analyse de la variance (ANOVA) a ensuite été réalisée afin d'observer d'éventuelles différences entre les moyennes des différentes périodes. Les résultats n'ont globalement pas révélé d'effets des conditions expérimentales (c.-à-d. des fluctuations de qualité) sur les données obtenues. Pour améliorer la démarche, une investigation s'est développée suivant deux axes : l'extraction de nouveaux indices (section 2) et la proposition d'un modèle empirique de détection automatique (section 3) pour observer les éventuelles réactions physiologiques liées à la présence de dégradations.

2 Traitement des signaux et extraction d'indicateurs

Dans le cadre de cette étude, des techniques d'analyse des signaux bruts (FC, AED, TCP, VSP) ont été développées pour permettre de dégager plus facilement des indicateurs de changement d'état ou « d'événements ». Par ailleurs, l'objectif d'une procédure aussi automatisée que possible amène à élaborer des algorithmes s'adaptant à de fortes variabilités interindividuelles et même intra-individuelles (sur la durée de l'expérience), là où les algorithmes standards demandent en général un ajustement manuel au cas par cas (de seuils, *etc.*) ou bien manquent de robustesse.

Dans l'exemple illustré en Figure 2, un détecteur a permis de transformer la courbe continue de l'AED en représentation parcimonieuse de réactions dont l'amorce est localisée temporellement et l'intensité quantifiée. Ce détecteur s'appuie sur l'identification de constantes individuelles de pentes en période de relaxation, estimées sur le long terme (méthode différente de [4] bien que d'esprit similaire).

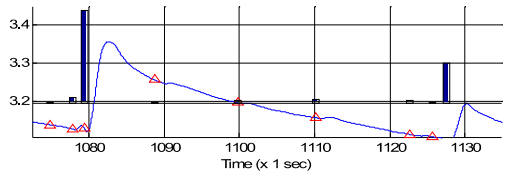


Figure 2 : Conductance électrodermale (courbe bleue) et indicateurs d'événements extraits (triangle : amorce réaction, barre verticale : intensité réaction)

La FC quant à elle est classiquement exploitée à travers sa variabilité temporelle et son observation en bandes de fréquences (hautes, basses et très basses fréquences). Ces indicateurs sont à relativement long terme au regard des événements que l'on cherche à identifier. C'est pourquoi des caractéristiques de l'onde de pouls (VSP), plus en amont, sont extraites pour être prises en compte dans un détecteur tel que décrit en section 3. La Figure 3 permet d'observer des variations de l'enveloppe d'amplitude (simple à évaluer) et de la forme d'onde de chaque battement (qui traduit l'élasticité des artères et reflète un effet du stress [5,6]). Ces caractéristiques apparaissent pour partie corrélées à d'autres signaux comme l'AED et la TCP et doivent permettre à la fois de consolider et affiner le détecteur.

Enfin, il importe que la FC elle-même et ses indicateurs dérivés participent efficacement à la détection de réactions aux stimuli et conditions de l'expérience. Or, les mesures recueillies pour certains sujets ont montré des anomalies fréquentes (battements manquants et/ou fausses détections). Un algorithme exploitant des statistiques locales de l'onde de pouls a donc été élaboré afin de fiabiliser la détection des battements cardiaques. Il permet par la même occasion d'identifier d'éventuels accidents (battements vraiment manquants, précoces ou bien tardifs), susceptibles d'être expliqués par des facteurs individuels et non expérimentaux, ou *a minima* d'être isolés. Comme illustré dans la Figure 3, l'algorithme fournit une détection robuste des pulsations cardiaques, complétée par le marquage et la caractérisation d'anomalies effectives. On peut alors choisir de les intégrer telles quelles ou bien de les « régulariser » avant calcul de la FC et indicateurs dérivés.

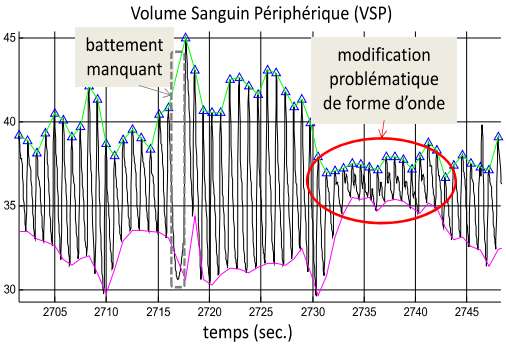


Figure 3 : Exemple d'onde de pouls (en noir) avec marqueurs de battements détectés (triangles bleus), enveloppe d'amplitude (min : magenta ; max : verte). La modification d'enveloppe et de forme d'onde cerclée de rouge est sujette à fausses détections ou détections manquantes avec des algorithmes standards

3 Détection automatique de la présence de dégradations

En contexte d'analyse de signaux multiples et de prise de décision, un modèle empirique de détection de l'influence de dégradations sur le pattern physiologique de l'individu a ensuite été élaboré en combinant

plusieurs signaux au sein d'un même modèle de Markov caché (HMM Hidden Markov Model) [7, 8].

Les paramètres d'entrée du modèle sont des vecteurs de signaux physiologiques : AED et indicateurs d'AED (section 2), TCP, FC (calculée en battements par minute) et/ou autres indicateurs associés (section 2). La phase d'apprentissage a permis d'estimer les densités de probabilité du modèle individuellement à partir de quatre contenus. Le test a été réalisé sur l'ensemble des signaux physiologiques mesurés durant la visualisation des cinq contenus de test par dix-sept individus (choisis pour la qualité de leurs mesures). La Figure 4 présente les résultats obtenus pour deux individus. Elle permet d'observer un certain nombre d'occurrences de détection pendant les périodes de dégradation mais aussi de nombreuses fausses alarmes.

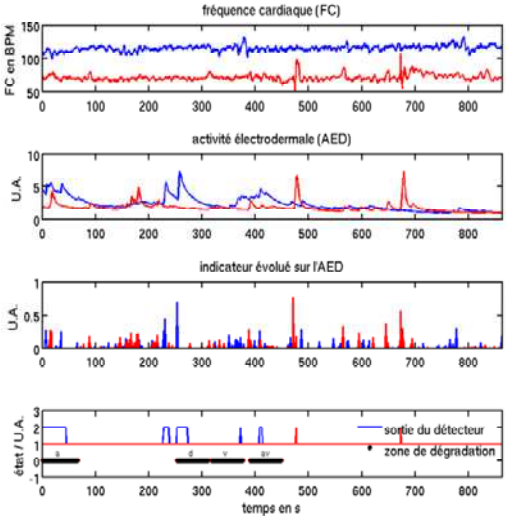


Figure 4 : Signaux d'entrée du modèle, état détecté en sortie du modèle et localisation des zones de dégradations

La présence de fausses alarmes pourrait être expliquée par une influence du contenu de la séquence (changement de plan, dynamique, intérêt, émotion induite, *etc*., caractérisés lors d'un questionnaire). Des images représentatives des différents niveaux de dynamique possibles au sein d'un même contenu sont apportées dans la Figure 5.

Figure 5 : Images extraites du contenu Documentaire. L'image de gauche illustre une dynamique faible (repos), celle de droite une dynamique élevée (combat)

La Figure 6 présente les variations du niveau de dynamique (en lien avec les mouvements de caméra) en parallèle des détections d'évènements obtenus sur l'ensemble des individus.

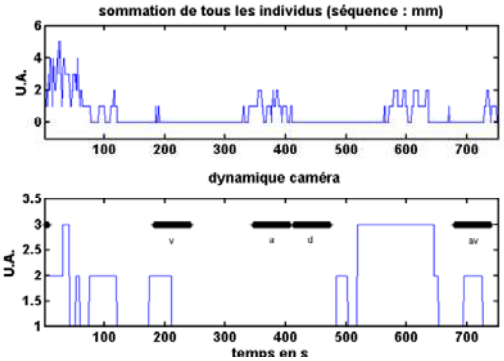


Figure 6 : Etat détecté et variation de la dynamique caméra (1 correspondant à une dynamique faible et 3 à une dynamique forte) pour le contenu Documentaire

La Figure 6 permet également de constater un effet du début du contenu pouvant être expliqué par un impact du changement d'activité (repos/visualisation), de découverte, de surprise ou encore d'anticipation (le début de chaque nouveau contenu était prévenu quelques secondes à l'avance). Cet effet lié au protocole a été observé pour chacun des cinq contenus de test.

Au-delà de l'influence du contenu ou du protocole, des facteurs propres à l'individu pourraient également être pris en compte comme des patterns spécifiques de réponse physiologique (individus labiles *vs*. stabiles) et éventuellement des facteurs psychologiques comme l'humeur (état émotionnel du moment), la fatigue, le stress, *etc.* Les différents facteurs d'influences, étudiés ou supposés, sur les mesures physiologiques sont schématisés dans la Figure 7 ci-après.

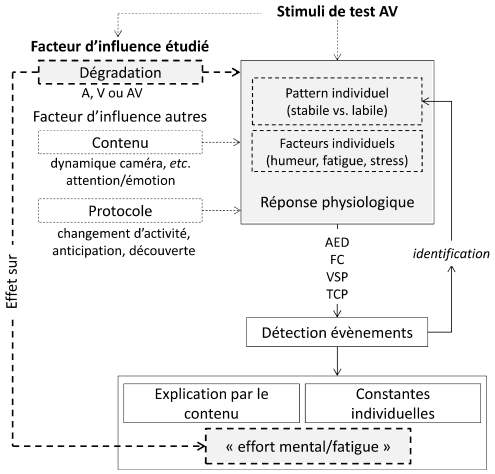


Figure 7 : Schéma présentant les différents facteurs d'influences des mesures physiologiques recueillies

4 Conclusions et perspectives

Dans le cadre d'une proposition de méthode alternative aux méthodes subjectives actuelles pour évaluer l'influence de la qualité audiovisuelle sur l'expérience du spectateur, un ensemble d'indicateurs physiologiques a été étudié. Différents traitements ont été réalisés dans l'optique de pouvoir détecter de manière automatique l'impact de dégradation de la qualité audiovisuelle à partir des signaux recueillis. Les résultats obtenus n'ont pas permis d'observer des réactions systématiques en présence de dégradation. La qualité de la détection effectuée doit être améliorée en envisageant par exemple un meilleur apprentissage et en tenant compte des différents facteurs d'influences de ces mesures (contenu, protocole, individu).

5 Références

[1] Wilson, G., & Sasse, A. (2000). Do Users Always Know What's Good For Them ? Utilising Physiological Responses to Assess Media Quality. Dans *Proceedings of HCI 2000 : People and Computers XIV - Usability or Else*! (p. 327-339), Sunderland, UK : Springer.

[2] Wilson, G., & Sasse, A. (2000). Investigating the impact of audio degradations on users : Subjective vs. Objective assessment methods. Dans *Proceedings of OZCHI 2000 : Interfacing Reality in the New Millennium*, 135-142.

[3] Lassalle, J., Gros, L., & Coppin, G. (2011). Combination of physiological and subjective measures to assess quality of experience for audiovisual technologies. Dans *Proceedings of the Third International Workshop on Quality of Multimedia Experience* (QoMEX), 13-18.

[4] Clarion, A. (2009). *Recherche d'indicateurs électrodermaux pour l'analyse de la charge mentale en conduit automobile* (Thèse de Doctorat, Université Lyon I, France). Récupérée du site thèse en ligne : http://tel.archives-ouvertes.fr.

[5] Kageyama, Y. Odagaki, M., & Hosaka, H. (2007). Wavelet Analysis for Quantification of Mental Stress Stage by Finger-tip Photo-plethysmography. Dans *Proceedings of IEEE, Engineering in Medicine and Biology Society* (EMBS), 1846-1849.

[6] Kim, K.H., Bang, S.W., & Kim, S.R. (2004). Emotion recognition system using short-term monitoring of physiological signals. *Medical and Biological Engineering and Computing*., 42(3), 419–427.

[7] Rabiner, L. R. (1989). A tutorial on Hidden Markov Models and selected applications in speech recognition. Dans *Proceedings of the IEEE*, 77(2), 257–286.

[8] Le Page, R. (2003). *Détection et analyse de l'onde P d'un électrocardiogramme : application au dépistage de la fibrillation auriculaire* (Thèse de doctorat, Université de Bretagne Occidentale, Brest, France).

ANNEXE 9-B

Expérimentation annexe

A. INTRODUCTION

Les résultats des expériences A et C ont montré qu'il était difficile d'observer un effet de la qualité sur les mesures psychophysiologiques. La visualisation seule des séquences audiovisuelles dont la qualité est fluctuante pourrait ne pas avoir été un stimulus suffisant pour déclencher des modifications importantes, ou tout au moins observables, des patterns de l'activité physiologique et oculaire.

Pour Wastell et Newman (1996[49]) une compréhension holistique du comportement humain repose sur l'étude conjointe de trois dimensions fondamentales : le comportement manifeste

[49] Voir REFERENCES

(performances), la physiologie et l'expérience subjective. Bernston *et al.* (1996[50]) ont constaté une décélération du rythme cardiaque (activation parasympathique) en l'absence de tâche explicite tel que cela peut être le cas lors d'une activité de visualisation de contenus audiovisuels. Ainsi, l'ajout d'une tâche explicite et parallèle à l'activité de visualisation pourrait constituer un protocole plus adapté à l'évaluation de l'impact de la qualité sur l'activité psychophysiologique du spectateur. Une hypothèse consisterait à croire que l'effort mental réalisé pour accomplir une tâche (comptage d'évènements survenant sur la modalité auditive ou visuelle par exemple) lors de la visualisation d'une séquence sera plus important en présence de dégradations que lors d'une séquence non dégradée. En effet, le partage attentionnel entre l'activité secondaire et principale pourrait être mis en difficulté par la présence de dégradations en considérant que le traitement de ces dernières nécessiterait des ressources attentionnelles plus importantes (voir Lang A. *et al.*, 2000[50]). L'augmentation de l'effort serait visible du point de vue des performances mais aussi à travers les mesures psychophysiologiques qui pourraient montrer de plus franches variations. Par ailleurs, il est généralement reconnu que plus les tâches (principales et secondaires) à réaliser sont proches, plus elles sont susceptibles d'interférer et de conduire à une diminution des performances pour l'une et/ou l'autre tâche (Chanquoy *et al.*, 2007[50]). Ainsi, des tâches à réaliser sur la modalité auditive (ou visuelle) lorsque le signal audio (ou vidéo) est dégradé pourraient être demandées au participant.

La présente expérience a été conçue sur la base de l'approche tripartite de Wastell et Newman (1996) pour étudier les effets de dégradations audio ou vidéo sur la *qualité d'expérience* du spectateur notamment concernant le *coût utilisateur* (performances, mesures subjectives et mesures psychophysiologiques) dans le cadre de l'ajout d'une tâche explicite (repérage de mots cibles ou d'actions/objets particuliers). Les dégradations étaient introduites lors de la présentation de séquences audiovisuelles 3D de courtes durées.

B. OBJECTIFS

Un premier objectif est de tester la pertinence du protocole proposé, l'ajout de la tâche doit permettre d'observer une influence de la qualité sur l'ensemble des mesures recueillies. Un second objectif consiste à étoffer le questionnaire subjectif notamment par l'ajout de critères plus spécifiques à l'évaluation de l'état émotionnel du participant et de la *qualité d'expérience* (QoE) du spectateur après la visualisation de chaque séquence audiovisuelle de test.

C. PARTICIPANTS

Trente participants (17 femmes et 13 hommes avec une moyenne d'âge de 31,5 ans), répartis en deux groupes de quinze personnes et non porteurs de lunettes pour faciliter

[50] Voir REFERENCES

l'enregistrement des mesures oculaires, ont participé à l'expérience. Ils étaient rémunérés pour leur contribution.

D. MATERIEL

Le matériel de test (configuration générale et technique, solution de synchronisation) et les techniques de recueil des données étaient strictement identiques à ceux de l'expérimentation C.

E. STIMULI

Quatre séquences AV ont été extraites de quatre des cinq contenus du corpus de l'expérimentation C à savoir *Documentaire*, *Opéra*, *Sport* et *Théâtre*. Une séquence d'environ une minute quinze était extraite de chaque contenu. Les séquences étaient présentées au format 3D (vidéo) stéréo (audio) full HD 1080p et préalablement encodées à 15 Mbps (AVC-x264, .avi). Les tâches de comptage d'évènements auditifs étaient toujours réalisées durant les séquences *Documentaire* et *Théâtre*, les séquences *Sport* et *Opéra* étaient dédiées au comptage d'évènements visuels.

Chaque séquence était visualisée deux fois, une fois sans dégradations et une fois avec une dégradation audio ou vidéo. A chaque présentation correspondait une tâche de comptage d'évènements auditifs (si la dégradation était audio) ou visuels (si la dégradation était vidéo). Pour une séquence donnée, deux tâches de comptage différentes étaient donc prévues. Par exemple, pour la séquence *Opéra*, une tâche de comptage d'évènements visuels était prévue pour la visualisation sans dégradations et une seconde tâche de comptage d'évènements visuels était prévue pour la présentation avec la dégradation vidéo. Ainsi, un participant ne réalisait jamais deux fois la même tâche bien qu'une séquence donnée était vue et entendue deux fois. Cela devait permettre d'éviter un effet d'apprentissage. Au total, huit différentes tâches ont été élaborées. Celles-ci ont été pensées pour être les plus équivalentes possibles (nature et difficulté : entre 7 et 14 évènements audio ou vidéo à repérer dans une séquence donnée) entre les différentes séquences. Le détail des tâches demandées est décrit ci-après. Les abréviations sont à lire comme suit : Tâche « T » 1 (T1) ou 2 (T2) à réaliser pour la séquence *Documentaire* (Doc), *Théâtre* (Th), *Sport* (Sp) ou *Opéra* (Op). Ainsi la tâche 1 correspondant à la séquence Documentaire sera notée T1_Doc.
Tâche « audio » ou comptage de mots cibles pour les séquences *Documentaire* et *Théâtre* :

- **T1_Doc** : mot cible « je » (× 12),
- **T2_ Doc** : mot cible « on » (× 8),
- **T1_ Th** : mot cible « vous » (× 10),
- **T2_ Th** : mot cible « je » (× 7).

Tâche « vidéo » ou comptage d'évènements visuels pour les séquences *Sport* et *Opéra* :

- **T1_ Sp** : évènement cible « balles passant au-dessus du filet » (× 14),

- **T2_ Sp** : évènement cible « changement de plan » (× 11),
- **T1_ Op** : évènement cible « contact physique entre les acteurs » (× 9),
- **T2_Op** : évènement cible « entrées/sorties des acteurs » (× 11).

Une illustration des tâches vidéo T2_Sp et T1_Op est apportée par les Figures 3 et 4 ci-après.

Fig. 3. Tâche « vidéo » T2_Sp pour laquelle les participants devaient compter le nombre de changements de plan durant la séquence.

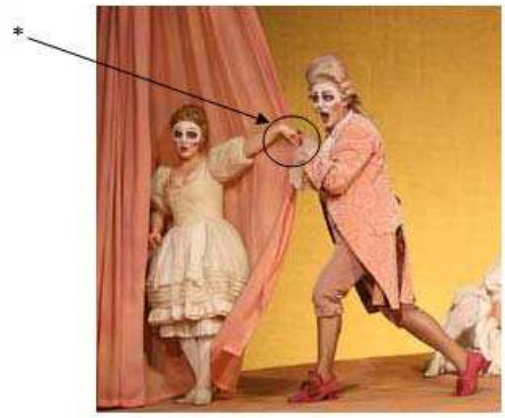


Fig. 4. Tâche « vidéo » T1_Op pour laquelle les participants devaient compter le nombre de contacts physique entre les acteurs (*) durant la séquence.

Deux types de dégradations étaient introduits : audio (perte de paquets, dégradation A-PP, chap. VII) et vidéo (diminution du débit variable entre 85 Kbps et 465 Kbps selon le contenu, dégradation V-DEB, chap. VII). La dégradation audio était toujours appliquée aux séquences dont la tâche était « audio » (Documentaire et Théâtre) tandis que la dégradation vidéo était toujours présentée avec les séquences pour lesquelles la tâche était « vidéo » (Sport et Opéra).

Deux groupes de participants visualisaient l'ensemble des séquences. Pour le premier, la dégradation était toujours introduite lors de « T1 » tandis que « T2 » ne présentait pas de dégradations. Le pattern inverse était présenté au second groupe de façon à contrebalancer un éventuel effet de la tâche. Ainsi, chaque condition (dégradation × séquence) était vue et

entendue par le panel de participants. Le Tableau I ci-après récapitule le plan d'expérience. Les huit conditions étaient présentées selon un ordre aléatoire différent pour chaque sujet.

Tableau 1. Plan d'expérience pour chaque groupe de participants (1 et 2) où VI correspond aux variables indépendantes (stimuli). La première ligne indique le type de tâche (Auditive : A ou Visuelle: V), la seconde ligne donne le nom de la séquence pour laquelle la tâche sera réalisée, la dernière ligne renseigne sur l'absence (Ø) ou la présence de dégradations audio (A) ou vidéo (V).

GROUPE 1

VI	**T1_Doc**	**T2_Doc**	**T1_Th**	**T2_Th**	**T1_Sp**	**T2_ Sp**	**T1_Op**	**T2_Op**
Tâche	A	A	A	A	V	V	V	V
Séquence	Doc.	Doc.	Théâtre	Théâtre	Sport	Sport	Opéra	Opéra
Dégradation	Ø	A	Ø	A	Ø	V	Ø	V

GROUPE 2

VI	**T1_Doc**	**T2_Doc**	**T1_Th**	**T2_Th**	**T1_Sp**	**T2_ Sp**	**T1_Op**	**T2_Op**
Tâche	A	A	A	A	V	V	V	V
Séquence	Doc.	Doc.	Théâtre	Théâtre	Sport	Sport	Opéra	Opéra
Dégradation	A	Ø	A	Ø	V	Ø	V	Ø

F. OBSERVABLES

F.1 PERFORMANCES

Les participants devaient, tout de suite après la visualisation d'une séquence donnée, reporter sur un questionnaire papier le nombre d'évènements auditifs ou visuels (performances) comptabilisé durant la visualisation de la séquence.

F.2 MESURES SUBJECTIVES

L'évaluation des niveaux perçus de qualité audiovisuelle (AV), vidéo (V) et audio (A) était identique à celle de l'expérimentation C. Dans cette expérience, un ensemble de critères subjectifs supplémentaires a également été soumis à l'évaluation des participants :

- **Évaluation des émotions ressenties** : quatre adjectifs étaient présentés sous la forme d'une échelle sémantique différentielle (valences opposées : positive et négative, situées à chaque extrémité de l'échelle) en neuf points (de - 4 à 4) (développée par Février, 2011[51]). La Figure 1 ci-après présente le questionnaire utilisé pour l'autoévaluation de l'état émotionnel du spectateur,

[51] Février, F. (2011). *Vers un modèle intégrateur "expérience-acceptation" : Rôle des affects et de caractéristiques personnelles et contextuelles dans la détermination des intentions d'usage d'un environnement numérique de travail* (thèse de doctorat, Université Rennes 2, France). Récupéré de http://tel.archives-ouvertes.fr/docs/00/60/83/35/PDF/theseFevrier.pdf

	-4	-3	-2	-1	0	1	2	3	4	
Frustré	o	o	o	o	o	o	o	o	o	**Satisfait**
Enervé	o	o	o	o	o	o	o	o	o	**Calme**
Tendu	o	o	o	o	o	o	o	o	o	**Détendu**
Mal à l'aise	o	o	o	o	o	o	o	o	o	**A l'aise**

Fig. 1. Echelles sémantiques différentielles proposées pour évaluer quatre adjectifs devant décrire les émotions ressenties durant la visualisation des séquences de test (d'après Février, 2011).

- **Évaluation de quatre autres dimensions de QoE** : effort mental, intérêt, compréhension et qualités hédoniques (sur la base des travaux d'Hassenzahl, Dieffenbach et Goritz, 2010[52]). Chaque dimension était évaluée à partir de trois questions lui étant spécifiques. Ces quatre critères sont regroupés, dans la suite du document, sous l'appellation « QoE » pour établir une distinction avec l'évaluation de l'état émotionnel et de la qualité perçue. La Figure 2 présente le questionnaire proposé.

[52]Hassenzahl, M., Dieffenbach, S. et Goritz, A. (2010). Needs, affect, and interactive products – Facets of user experience. *Interacting with Computers*, *22*(5), 353–362.

		1	2	3	4	5	6	7	8	9	10
1	Il faut être concentré pour comprendre ce type de séquence audiovisuelle en 3D	○	○	○	○	○	○	○	○	○	○
2	Cette séquence audiovisuelle était stimulante	○	○	○	○	○	○	○	○	○	○
3	J'estime pouvoir expliquer le contenu de la séquence à quelqu'un	○	○	○	○	○	○	○	○	○	○
4	La visualisation de cette séquence m'a demandé un effort mental important	○	○	○	○	○	○	○	○	○	○
5	J'ai été intéressé(e) par cette séquence audiovisuelle	○	○	○	○	○	○	○	○	○	○
6	J'ai éprouvé(e) de la difficulté pour accomplir la tâche qui m'a été demandée	○	○	○	○	○	○	○	○	○	○
7	J'ai bien aimé le contenu de cette séquence audiovisuelle	○	○	○	○	○	○	○	○	○	○
8	Il était facile de comprendre le contenu de cette séquence audiovisuelle	○	○	○	○	○	○	○	○	○	○
9	Cette séquence était originale	○	○	○	○	○	○	○	○	○	○
10	Cette séquence était agréable à regarder	○	○	○	○	○	○	○	○	○	○
11	J'estime avoir compris les informations contenues dans cette séquence audiovisuelle	○	○	○	○	○	○	○	○	○	○
12	J'ai tellement apprécié cette séquence que j'aimerais en savoir plus sur le sujet	○	○	○	○	○	○	○	○	○	○

Fig. 2. Questionnaire d'évaluation, des différentes dimensions (Effort mental : questions 1, 4, 6, Intérêt : questions 5, 7, 12, Compréhension : questions 3, 8, 11 et Qualités hédoniques : questions 2, 9, 10) de la QoE, composé de douze affirmations. Les participants devaient évaluer leur degré d'accord avec chaque affirmation à l'aide d'une échelle en 10 points où 1 = pas du tout d'accord et 10= tout à fait d'accord.

F.3 MESURES PHYSIOLOGIQUES ET OCULAIRES

Les mesures physiologiques et oculaires étaient identiques à celles de l'expérimentation C.

G. PROTOCOLE

Le protocole appliqué pour la préparation et l'installation des capteurs physiologiques et oculaires était strictement identique à celui de l'expérimentation C. Aucune baseline n'était enregistrée, les différences seules entre conditions dégradées et non dégradées étant étudiées. Chaque participant visualisait un total de huit séquences AV (4 séquences × 2 dégradations). Entre chaque séquence, les participants disposaient de cinq minutes pour compléter les questionnaires proposés et noter le résultat obtenu à la tâche de comptage. Au total un participant reportait huit scores, un pour chaque séquence audiovisuelle visualisée et donc pour chaque tâche réalisée. Au terme des cinq minutes, une nouvelle séquence débutait. Afin

d'éviter tout effet de surprise, le participant était prévenu une minute puis six secondes avant le début de chaque nouvelle séquence. Pour chaque participant, l'ordre de passation des séquences était présenté de manière aléatoire. Il leur était également demandé de rester attentif à la tâche de visualisation et de ne pas tricher lors de la tâche de comptage (comptage mental sans support papier). Avant la présentation des conditions expérimentales, une phase d'apprentissage devait permettre au participant de se familiariser avec les tâches à réaliser durant le test. La phase d'apprentissage se composait de cinq séquences de trente secondes. La durée totale de passation était d'environ 1 heure 30.

H. HYPOTHESES

Dans cette expérimentation, les influences du niveau de qualité sur les performances, l'expérience subjective et l'activité psychophysiologique (physiologique et oculaire) ont été étudiées dans un contexte où une tâche explicite était demandée au spectateur en plus de l'activité de visualisation. Il était attendu que la présence de dégradations diminue les performances obtenues aux tâches de comptage, les scores moyens obtenus à l'évaluation des niveaux de qualité audio (MOSA), vidéo (MOSV) et audiovisuelle (MOSAV) ainsi que la *qualité d'expérience* (évaluation des dimensions de QoE). La présence de dégradations devait aussi entraîner une expérience émotionnelle plus négative. Il était également attendu que l'effort mental induit par la tâche réalisée en présence de dégradations soit reflété par les mesures psychophysiologiques. Les hypothèses suivantes ont été formulées :

- **H0** : la présence de dégradations diminuerait les performances obtenues aux tâches de comptage par rapport aux séquences sans dégradations,
- **H1** : la présence de dégradations serait reflétée par les évaluations subjectives (qualité, émotion et QoE),
- **H2** : la réalisation d'une tâche de comptage d'évènements auditifs ou visuels en présence de dégradations survenant sur la même modalité pourrait induire un effort mental puis éventuellement de fatigue observables à travers les mesures physiologiques et oculaires.

I. RESULTATS

Dans un souci de concision, seuls les effets significatifs sont présentés. Les figures ci-dessous présentent des intervalles de confiance à 95%.

I.1 PERFORMANCES

La Figure 5 ci-dessous présente le pourcentage de détection des évènements visuels ou auditifs comptabilisés (une performance de 100% correspond au nombre exact d'évènements visuels ou auditifs à comptabiliser durant la scène). La figure permet deux constats : les performances tendent à diminuer en présence de dégradations et le type de tâche (propre à une séquence) semble avoir influencé les résultats. Par exemple, la tâche 2 de la séquence *Opéra* a

entraîné un grand nombre de fausses détections (les participants ont rapporté un nombre plus élevé d'évènements que le nombre réel d'évènements présents dans la séquence). Ces observations sont confirmées par une ANOVA considérant les variables indépendantes « Séquence », « Tâche » et « Présence de dégradations » réalisée sur la variable dépendante « Performances ». Les résultats ont révélé un effet de la dégradation ($F(1,14) = 15,55$, $p < 0,05$) et de l'interaction des variables « Séquence » et « Tâche » ($F(3,42) = 22,03$, $p < 0,001$). Les résultats ont également indiqué que la tâche 1 de la séquence *Documentaire* a entraîné des performances significativement moins bonnes que celles de la tâche 2 réalisée pour la même séquence.

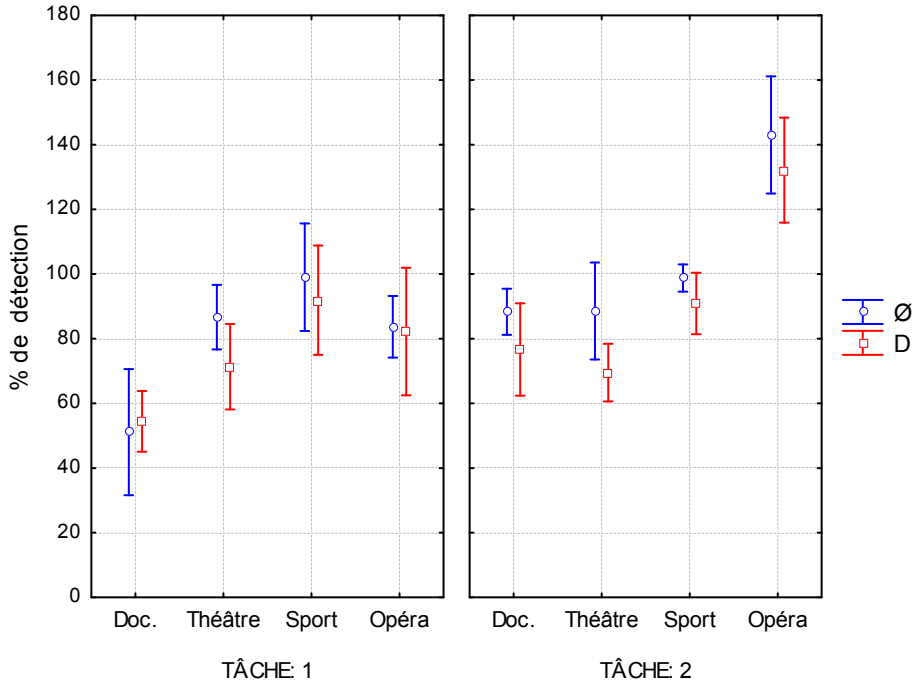


Fig. 5. Performances moyennes obtenues aux différentes tâches de comptage lors de la visualisation des séquences avec ou sans dégradations.

Conclusions Performances :

Il était attendu que la présence de dégradations diminue les performances obtenues aux tâches de comptage par rapport aux séquences sans dégradations (H0). Les résultats obtenus ont permis de confirmer ce postulat en révélant une influence de la qualité sur les performances des participants. Précisément, celles-ci étaient moins bonnes en présence de dégradations. Par ailleurs, indépendamment d'un effet de la dégradation, le type de tâche a également influencé les scores. La tâche 1 de la séquence *Documentaire* (mot cible « je ») et la tâche 2 de la séquence *Opéra* (évènement cible : entrée/sortie des acteurs) ont été les plus difficiles du corpus (performances les plus faibles ou erreurs les plus nombreuses). Le nombre important de fausses détections observé pour *Opéra*-T2 pourrait signifier que la tâche a été mal comprise ou mal interprétée par les participants.

I.2 MESURES SUBJECTIVES

Une seconde hypothèse (H1) consistait à croire que la présence de dégradations diminue la *qualité d'expérience* (qualité perçue, état émotionnel et autres dimensions de QoE) par rapport aux séquences sans dégradations.

I.2.1 QUALITÉ

La Figure 6 ci-dessous présente les scores moyens obtenus pour l'évaluation de la qualité audiovisuelle (MOSAV), vidéo (MOSV) et audio (MOSA). Pour rappel, la dégradation appliquée aux séquences *Documentaire* et *Théâtre* était audio tandis que la dégradation appliquée aux séquences *Sport* et *Opéra* était vidéo.

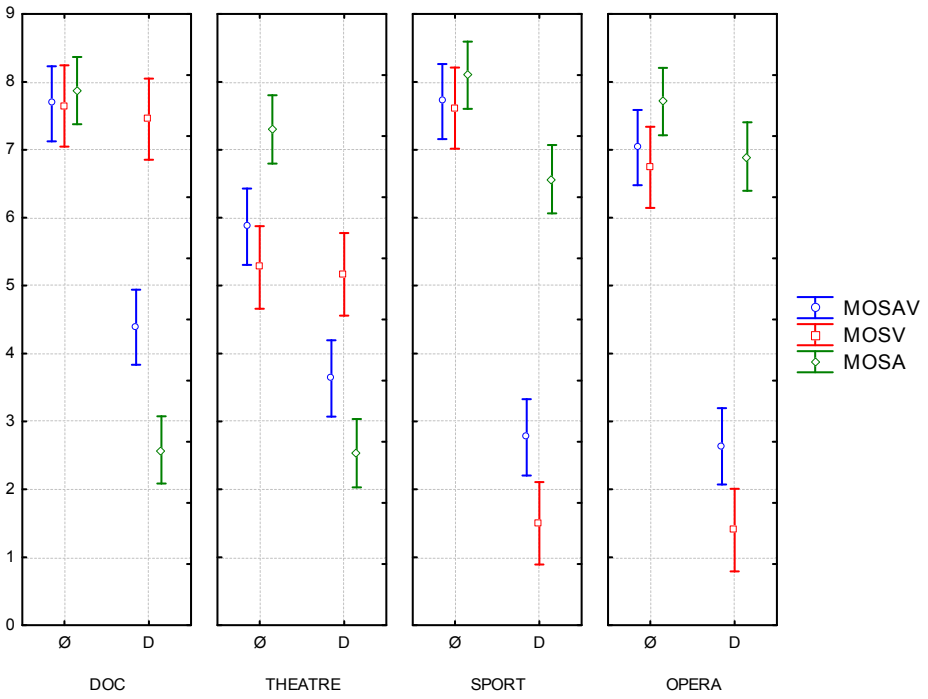


Fig. 6. MOSAV, MOSV et MOSA obtenues pour les conditions avec dégradation (D) ou sans dégradations (Ø) pour chacune des séquences visualisées : Documentaire (Doc.), Théâtre, Sport et Opéra, toutes tâches confondues.

La figure indique que MOSAV diminuait systématiquement en présence de la dégradation audio (Documentaire et Théâtre) ou vidéo (Sport et Opéra). MOSA diminuait en présence de la dégradation audio, de la même manière, MOSV diminuait lorsque la dégradation vidéo était présentée. La figure montre également que la qualité vidéo n'a pas été influencée par la présence de la dégradation audio (Documentaire et Théâtre). En revanche, la qualité audio a été impactée par la dégradation vidéo (Sport et Opéra).

Une ANOVA considérant les variables indépendantes « Séquence » (Documentaire, Théâtre, Sport, Opéra), « Tâche » (T1 et T2) et « Présence de dégradations » (dégradée, non dégradée) et considérant la variable aléatoire « Participant » a été conduite sur les variables

dépendantes « Qualité AV», « Qualité V» et « Qualité A». Aucun effet de la tâche n'a été mis en évidence par l'analyse. Les résultats significatifs obtenus sont présentés ci-dessous.

Qualité Audiovisuelle : un effet significatif de la séquence (F (3,199) =12,27, $p < 0{,}001$), de la dégradation (F (1, 199) =476,49, $p < 0{,}001$) et de l'interaction (F (3,199) = 12,45, $p < 0{,}001$) entre « Séquence » et « Présence de dégradations » a été observé. La qualité audiovisuelle a été jugée comme satisfaisante en présence de la dégradation audio (Documentaire et Théâtre) et comme médiocre en présence de la dégradation vidéo (Sport et Théâtre). La qualité audiovisuelle était notée comme satisfaisante ou bonne en l'absence de dégradations.

Qualité Audio : un effet significatif de la séquence (F (3,199) =64,66, $p < 0{,}001$), de la dégradation (F (1,199) =374,39, $p < 0{,}001$) et de l'interaction (F (3,199) = 48,81, $p < 0{,}001$) entre « Séquence » et « Présence de dégradations » a été observé. La qualité audio a été jugée comme médiocre en présence de la dégradation audio (Documentaire et Théâtre) et comme bonne en présence de la dégradation vidéo. La qualité audio était notée comme bonne ou excellente en l'absence de dégradations.

Qualité Vidéo : un effet significatif de la séquence (F (3,199) =73,09, $p < 0{,}001$), de la dégradation (F (1,199) =262,39, $p < 0{,}001$) et de l'interaction (F (3,199) = 79,59, $p < 0{,}001$) entre « Séquence » et « Présence de dégradations » a été observé. La qualité vidéo a été jugée comme mauvaise en présence de la dégradation vidéo (Sport et Théâtre) et comme bonne ou satisfaisante en présence de la dégradation audio. La qualité vidéo était notée comme satisfaisante ou bonne en l'absence de dégradations. La qualité vidéo de la séquence *Théâtre* a été jugée comme étant moins bonne que celle des autres séquences.

Conclusions Qualité :

Globalement, les dégradations audio et vidéo ont entraîné une diminution des notes de qualité audio, vidéo et audiovisuelle. Les notes de qualité audiovisuelle semblent plus fortement influencées par les notes de qualité vidéo comme en témoigne notamment les notes de QAV obtenues en présence de la dégradation vidéo ou celles attribuées à la séquence *Théâtre* présentée sans dégradations. La note moyenne de qualité vidéo obtenue pour *Théâtre* reflète probablement la qualité 3D native du contenu moins bonne pour cette séquence (voir expérience C). Par ailleurs, les notes de qualité vidéo étaient également moins bonnes en présence de la dégradation vidéo que les notes de qualité audio en présence de la dégradation audio. En d'autres termes, les dégradations vidéo dégradent plus fortement la qualité perçue que les dégradations audio. Ce résultat pourrait témoigner d'un écart entre les seuils de dégradation choisis. Enfin, les notes de qualité vidéo n'ont pas été impactées en présence de la dégradation audio (séquences Documentaire et Théâtre), en revanche, les notes de qualité audio ont diminué en présence de la dégradation vidéo (séquences Sport et Opéra). L'impact prédominant de la qualité vidéo sur la qualité audiovisuelle et audio va dans le sens des études

menées par Beerends et Caluwe (1999[53]) et la « Commission 12 » de l'UIT (COM12-61-E, 1998[53]) (voir chap. II).

I.2.2 EMOTIONS

La Figure 7 ci-dessous présente les moyennes obtenues pour chaque émotion évaluée : frustration (frustré-satisfait), énervement (énervé-calme), tension (tendu-détendu) et confort (mal à l'aise-à l'aise).

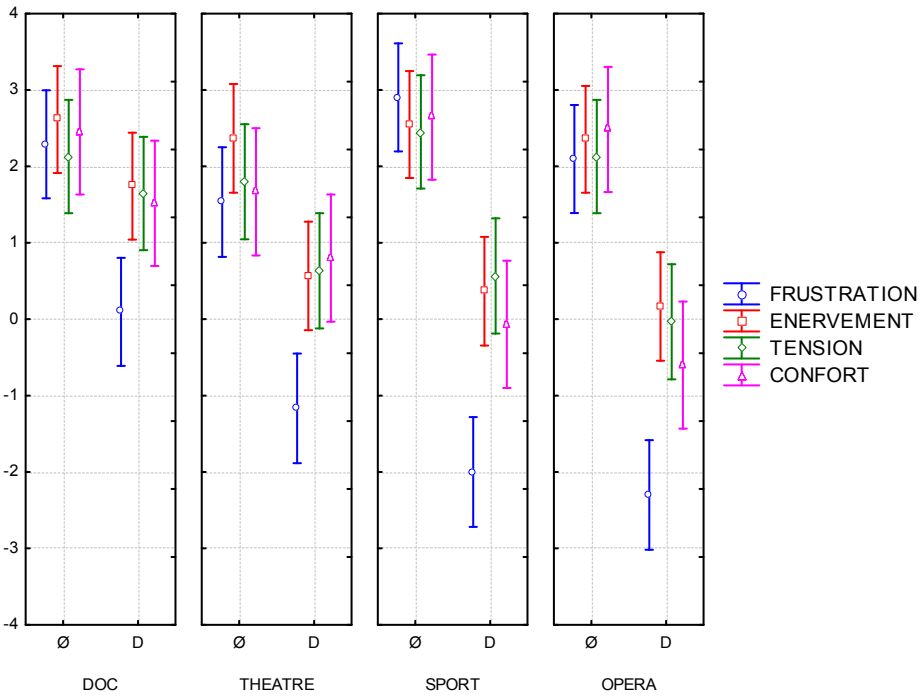


Fig. 7. Niveaux moyens obtenus pour chaque état émotionnel évalué après la visualisation des séquences extraites de chaque contenu (Documentaire –DOC.-, Théâtre, Sport et Opéra). Ø représente l'absence de dégradations et D la présence de dégradations, toutes tâches confondues. Plus le niveau est négatif, plus le participant déclarait être frustré, énervé, tendu ou mal à l'aise.

La figure indique que globalement l'état émotionnel du participant a été impacté par la présence de dégradations et notamment de la dégradation vidéo. Pour chaque état émotionnel « Frustration», « Enervement», « Tension » et « Confort », une ANOVA a été réalisée en prenant en compte les variables indépendantes « Séquence », « Tâche » et « Présence de dégradations » et considérant la variable aléatoire « Participant ». Aucun effet de la tâche n'a été mis en évidence par l'analyse. Les résultats significatifs obtenus sont présentés ci-dessous.

Frustration : un effet significatif de la séquence ($F(3,199) = 6{,}59$, $p < 0{,}001$), de la dégradation ($F(1,199) = 259{,}02$, $p < 0{,}001$) et de l'interaction ($F(3,199) = 9{,}01$, $p < 0{,}001$)

[53] Voir REFERENCES

entre « Séquence » et « Présence de dégradations » a été observé. Globalement, le sentiment de frustration augmentait en présence de dégradations.

Enervement : un effet significatif de la séquence ($F(3,199) = 3,26$, $p < 0,05$) et de la dégradation ($F(1,199) = 66,31$, $p < 0,001$) a été trouvé. Globalement, le sentiment d'énervement augmentait en présence de dégradations.

Tension : un effet significatif de la dégradation ($F(1,199) = 42,57$, $p < 0,001$) et de l'interaction ($F(3,199) = 3,03$, $p < 0,05$) entre « Séquence » et « Présence de dégradations » a été révélé. Globalement, le sentiment de tension augmentait en présence de dégradations.

Confort : un effet significatif de la séquence ($F(3,199) = 3,84$, $p < 0,05$), de la dégradation ($F(1,199) = 76,20$, $p < 0,001$) et de l'interaction ($F(3,199) = 7,13$, $p < 0,001$) entre « Séquence » et « Présence de dégradations » a été mis en évidence. Globalement, le sentiment de confort diminuait en présence de dégradations.

Conclusions Emotions :

Globalement, l'expérience émotionnelle des participants était plus « négative » en présence de dégradations (les participants ont rapporté être plus frustrés, énervés, tendus et moins à l'aise) par rapport à la condition où la dégradation était absente. De manière générale, la dégradation vidéo (Sport et Opéra) a plus fortement diminué la valence des émotions évaluées que la dégradation audio. Ce constat rejoint celui de l'évaluation de qualité.

L'état émotionnel ayant le plus fortement « réagi » à la présence de dégradations (plus grand pouvoir de discrimination), tel que subjectivement évalué ici, était celui concernant le sentiment de frustration qui augmentait fortement (jusqu'à environ - 5 points) en présence de dégradations.

Concernant l'effet de la dégradation audio, la séquence *Théâtre* a été à l'origine d'un sentiment de frustration plus important que la séquence *Documentaire*. Deux explications peuvent être apportées ici : soit le niveau de frustration reflète le niveau de qualité 3D, moins bon que pour les autres séquences, soit le niveau de frustration s'explique par une perte d'intelligibilité plus importante pour cette séquence (voir Expérimentation C : présence de dialogues diégétiques entre plusieurs interlocuteurs, la séquence *Documentaire* présentait principalement des sons de parole non diégétiques -voix *off*- ou un monologue). Cette dernière supposition pourrait permettre d'expliquer le niveau de performance plus faible obtenu pour l'une des tâches à accomplir lors de la visualisation de *Théâtre*. L'effet de la perte d'intelligibilité sur le niveau de frustration pourrait être confirmé par l'augmentation de l'effort mental ou une diminution de la compréhension lors de l'évaluation des dimensions de QoE pour cette séquence. Dans tous les cas, ce résultat révèle un effet du contenu.

En résumé, les résultats relatifs à l'état émotionnel des participants ont montré un bon niveau de discrimination entre la présence ou l'absence de dégradations notamment l'évaluation de l'état de frustration. Ils ont également indiqué un effet plus préjudiciable de la dégradation vidéo sur le niveau de frustration, de nervosité, de tension et de confort.

I.2.3 QoE

L'évaluation de la QoE a été réalisée à l'aide de trois questions relatives à chaque dimension à savoir l'effort mental ressenti, l'intérêt, la compréhension et la qualité hédonique de la séquence (stimulante, agréable, originale). Une moyenne a ensuite été calculée pour chaque dimension sur la base des réponses obtenues à chacune des trois questions qui la composaient. La Figure 8 ci-après présente les résultats obtenus.

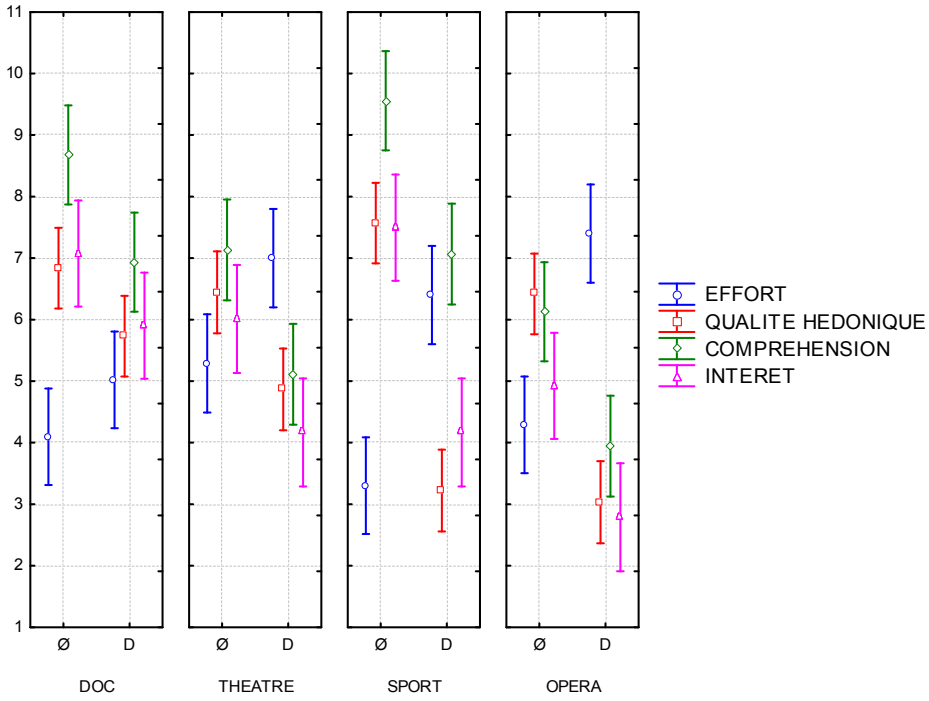


Fig. 8. Niveaux moyens obtenus pour chaque dimension de QoE évaluée après la visualisation des séquences extraites de chaque contenu (Documentaire –DOC.-, Théâtre, Sport et Opéra), toutes tâches confondues. Ø représente l'absence de dégradation sur la séquence et D la présence de dégradations. Les participants devaient évaluer leur degré d'accord à l'aide d'une échelle en 10 points où 1 = pas du tout d'accord et 10= tout à fait d'accord.

La figure indique que globalement la QoE a été impacté par la présence de dégradations, c'est-à-dire que l'effort mental ressenti augmentait tandis que la qualité hédonique, le niveau de compréhension et le niveau d'intérêt diminuaient et ce, indépendamment du type de séquence. Pour chaque dimension « Effort », « Qualité hédonique », « Compréhension » et « Intérêt », une ANOVA a été réalisée en prenant en compte les variables indépendantes « Séquence », « Tâche » et « Présence de dégradations » et considérant la variable aléatoire « Participant ». Aucun effet de la tâche n'a été mis en évidence par l'analyse. Les résultats significatifs obtenus sont présentés ci-dessous.

Effort : un effet significatif de la séquence (F (3,199) = 10,09, p < 0,001), de la dégradation (F (1,199) = 88,96, p < 0,001) et de l'interaction (F (3,199) = 5,48, p < 0,01) entre « Séquence » et « Présence de dégradations » a été trouvé. Il est également intéressant de noter que le comptage d'évènements auditifs durant la séquence *Théâtre* (avec ou sans la

dégradation audio) a demandé un effort mental plus important par rapport à la séquence *Documentaire*. Les résultats montrent également que l'effort mental ressenti était plus important lors de la séquence *Opéra* (avec ou sans la dégradation vidéo) par rapport à la séquence *Sport*.

Qualité Hédonique : un effet significatif de la séquence (F (3,199) = 8,62, p < 0,001), de la dégradation (F (1,199) = 141,63, p < 0,001) et de l'interaction (F (3,199) = 11,99, p < 0,001) entre « Séquence » et « Présence de dégradations » a été observé. La Figure 8 permet d'observer que toutes séquences confondues, la qualité hédonique de la séquence diminuait de manière significative en présence de dégradations, c'est-à-dire que les participants considéraient les séquences comme moins stimulantes, moins originales et moins agréables à regarder.

Compréhension : un effet significatif de la séquence (F (3,199) = 35,53, p < 0,001) et de la dégradation (F (1,199) = 68,92, p < 0,001) a été mis en évidence. De manière générale, le niveau de compréhension diminuait, toutes séquences confondues, en présence de dégradations. La séquence *Opéra*, présentée avec ou sans dégradations, était la moins comprise de l'ensemble du corpus de test.

Intérêt : un effet significatif de la séquence (F (3,199) = 15,16, p < 0,001) et la dégradation (F (1,199) = 53,24, p < 0,001) a été trouvé. Le niveau d'intérêt diminuait significativement en présence de dégradations, et ce, toutes séquences confondues. La séquence *Sport* a été jugée comme étant la plus intéressante lorsqu'aucune dégradation n'était appliquée. La séquence *Opéra* a été considérée comme la moins intéressante que la dégradation soit présente ou non.

Conclusions QoE :

Les dimensions évaluées, dans l'intention d'obtenir plus d'informations sur la *qualité d'expérience* du spectateur, ont montré un bon niveau de discrimination. De manière générale, les dégradations ont diminué la *qualité d'expérience* du spectateur.

L'effort mental ressenti par les participants était plus important lorsqu'une tâche devait être réalisée durant des séquences dégradées. Pour la tâche de comptage d'évènements auditifs (Documentaire et Théâtre), l'effort mental ressenti était plus important pour *Théâtre* que pour *Documentaire* en présence ou non de la dégradation audio. En parallèle, le niveau de compréhension était également plus faible pour *Théâtre* que pour *Documentaire*. Ce constat pourrait indiquer que la tâche de comptage (repérer des mots cibles) a interféré avec la compréhension de la séquence *Théâtre* fortement « verbale ». Ainsi, le repérage de mots au sein de séquences fortement verbales entraînerait, de fait, une baisse de la compréhension, l'individu étant plus amplement concentré sur la tâche de comptage que sur la compréhension du discours. Cet effet serait d'autant plus fort en présence de la dégradation audio qui entraverait l'intelligibilité du contenu. Cela tend à confirmer les résultats obtenus sur le sentiment de frustration plus important lors de cette séquence en présence de la dégradation audio.

Les participants ont rapporté que la séquence *Sport* était la plus stimulante, la plus agréable et la plus originale en l'absence de dégradations. Ce résultat peut être, en partie, attribué à un effet positif de la présentation 3D. Les participants ont effet rapporté une valeur ajoutée de la 3D pour cette séquence en proposant une expérience d'immersion plus importante que celles des autres séquences. La présence de dégradations diminuait, toutes séquences confondues, le niveau de qualité hédonique, c'est-à-dire que les séquences étaient jugées comme moins stimulantes, originales et agréables notamment lors de la dégradation vidéo. Il est intéressant de noter que ces aspects précisent la note de qualité obtenue.

Le niveau moyen de compréhension diminuait en présence de dégradations toutes séquences confondues. Les niveaux les plus faibles de compréhension ont été attribués aux séquences *Opéra* puis *Théâtre* en présence ou non de la dégradation audio. Ce résultat confirme ceux obtenus lors de l'expérimentation C présentant les contenus 3D en entier.

Le niveau moyen d'intérêt des séquences diminuait en présence de dégradations, toutes séquences confondues. La séquence *Opéra* a été jugée avec le niveau d'intérêt le plus faible en présence ou non de la dégradation vidéo. Ce résultat confirme le rejet déjà observé lors de l'expérimentation BI et C.

En conclusion, la présence de dégradations a été préjudiciable à la qualité perçue, à l'état émotionnel des participants ainsi qu'à la perception des autres dimensions de QoE. Ces résultats permettent de confirmer H1.

I.3 MESURES PSYCHOPHYSIOLOGIQUES

Les analyses ont été conduites à partir des mesures physiologiques et oculaires de vingt-neuf participants.

Une dernière hypothèse (H2) consistait à croire que la réalisation d'une tâche de comptage d'évènements auditifs ou visuels durant la visualisation d'une séquence présentant une dégradation sur la même modalité devrait induire un effort mental puis éventuellement de fatigue observables à travers les mesures physiologiques et oculaires. Ainsi, une série d'ANOVAs considérant les variables indépendantes « Séquence », « Tâche » et « Présence de dégradations » ainsi que la variable aléatoire « Participant » a été conduite sur chaque variable dépendante physiologique : « VSP », « AED », « TCP », « FC » et oculaire : « Ebdur », « EBfreq», « DP», « PERCLOS».

Les résultats ont révélé un effet de la séquence sur les indicateurs DP (F (3,189) = 4,60, $p < 0,01$) et EBfreq (F (3,189) = 4,49, $p < 0,01$), de l'interaction des deux variables « Séquence » et « Présence de dégradations » sur l'indicateur SAC (F (3,189) = 2,86, $p < 0,05$) et de l'interaction des trois variables indépendantes « Séquence », « Tâche » et « Présence de dégradations » sur EBfreq (F (3,189) = 2,8, $p < 0,05$). Les Figures 9, 10 et 11 présentent ces effets.

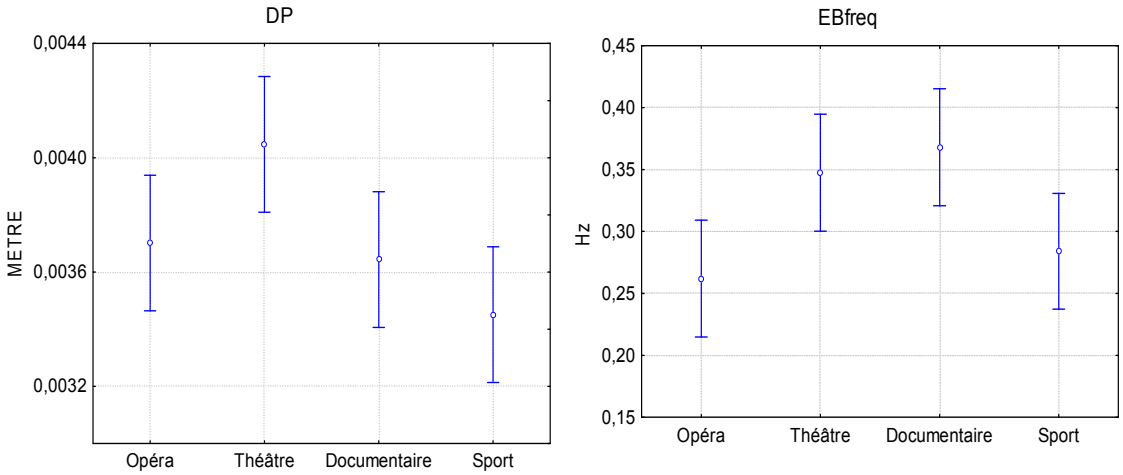


Fig. 9. Moyennes obtenues pour les indicateurs DP et EBfreq pour chaque séquence extraite des contenus Documentaire (Doc.), Théâtre, Sport et Opéra.

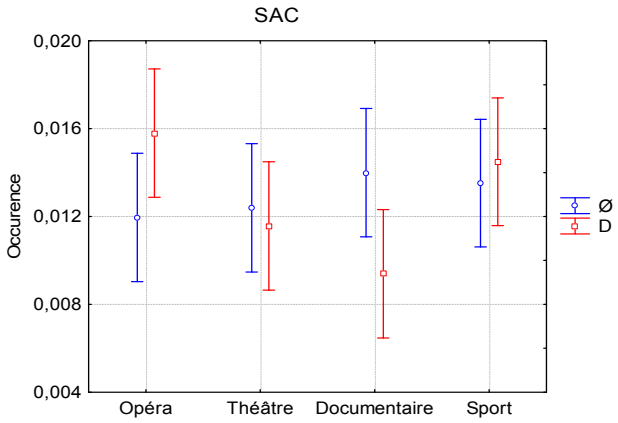


Fig. 10. Moyennes obtenues pour l'indicateur SAC durant la visualisation de chaque séquence extraite des contenus Documentaire (Doc.), Théâtre, Sport et Opéra présentées avec (D) ou sans (Ø) dégradation.

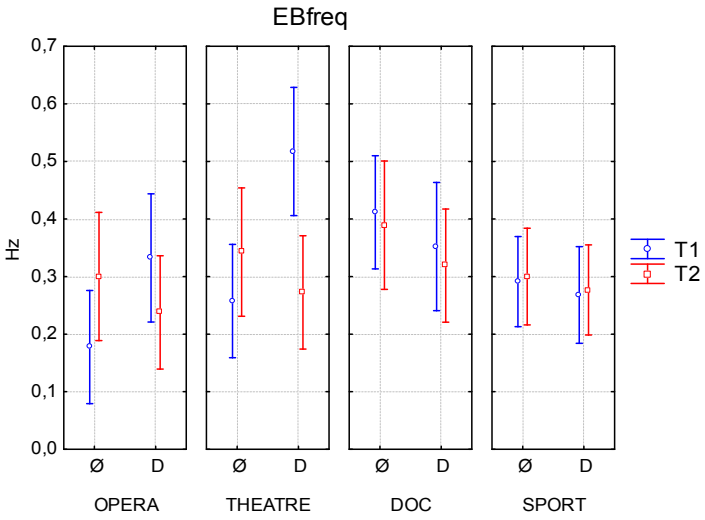


Fig. 11. Moyennes obtenues pour l'indicateur EBfreq lors de la réalisation des tâches (T1 ou T2) durant la visualisation de chaque séquence extraite des contenus Documentaire (Doc.), Théâtre, Sport et Opéra présentées avec (D) ou sans (Ø) dégradation.

Plus précisément, un test *post-hoc* HSD de Tukey a indiqué une différence significative entre les séquences *Sport* et *Théâtre* pour l'indicateur DP (avec $p < 0,01$) et entre les séquences *Documentaire* et *Opéra* pour l'indicateur EBfreq (avec $p < 0,01$). La taille du diamètre pupillaire était plus grande pour la séquence *Théâtre* que pour la séquence *Sport*. La Figure 9 indique une diminution de l'indicateur EBfreq lors de la séquence *Opéra* par rapport à la séquence *Documentaire*.

Un test *post-hoc* HSD de Tukey a également révélé une différence significative pour l'indicateur SAC entre les séquences *Opéra* et *Documentaire* (avec $p < 0,05$) lorsque celles-ci étaient dégradées. Une augmentation de SAC peut être observée pour la séquence *Opéra* par rapport à la séquence *Documentaire*.

Enfin, une différence significative pour l'indicateur EBfreq a également été trouvée entre la condition dégradée et non dégradée (avec $p < 0,05$) lors de la réalisation de la tâche 1 (T1) durant la séquence *Théâtre.* La réalisation de cette tâche lorsque la séquence était dégradée (audio) a été à l'origine d'une augmentation des EBfreq. Une augmentation, non significative, peut également être constatée pour la séquence Opéra (fig. 11). Globalement, EBfreq tendait à diminuer en présence de dégradations.

Les résultats ont aussi révélé un effet de l'interaction des variables indépendantes « Séquence », « Tâche » et « Présence de dégradations » sur l'AED : $F(3,189) = 3,08$, $p < 0,05$). La Figure 12 permet de constater que l'AED augmentait durant la réalisation de la tâche (généralement T2) notamment concernant la séquence *Théâtre*.

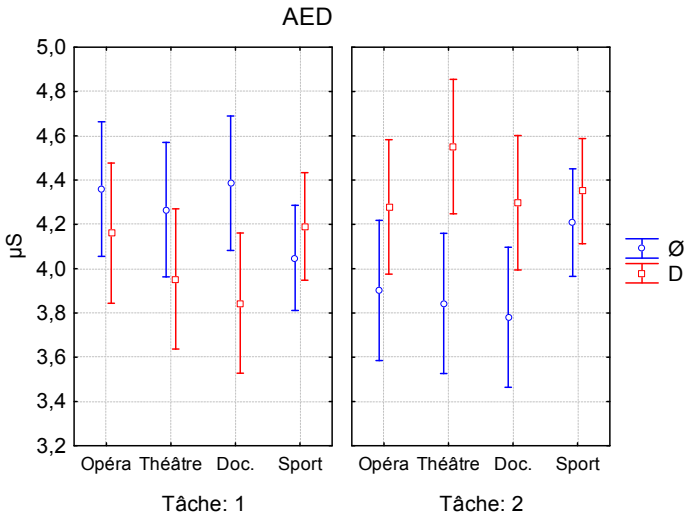


Fig. 12. Moyennes obtenues pour l'indicateur d'AED en présence (D) ou en absence (Ø) de dégradation.

Conclusions mesures psychophysiologiques :

Les résultats obtenus ont révélé un effet de la séquence sur le diamètre pupillaire et la fréquence de fermeture de l'œil. L'influence de la séquence sur la taille de la pupille est probablement liée au niveau de luminosité : le niveau le plus élevé correspondait à la séquence *Sport* tandis que la séquence *Théâtre* était caractérisée par le niveau le plus faible

(voir expérimentation C). La constriction observée pour *Sport* et la dilation observée pour *Théâtre* reflèteraient donc ces différences (réflexe photo-moteur). Ce résultat confirme la sensibilité de cet indice au niveau de luminosité difficilement contrôlable dans le cadre de visualisation de scènes n'étant pas des contenus de laboratoires.

La fréquence de clignements de l'œil a également été influencée par le type de séquence : les participants ont moins cligné des yeux durant *Opéra*. Cette diminution pourrait traduire une activation physiologique accrue durant cette séquence. Rappelons qu'*Opéra* a été jugé avec un niveau d'intérêt, de compréhension et de qualité hédonique plus faible que pour les autres séquences.

Une augmentation du nombre de saccades durant la séquence *Opéra,* comparativement au contenu *Documentaire,* lorsque la dégradation vidéo était présente a également été constatée. Sans dégradations, le nombre de saccade ne différait pas significativement d'une séquence à l'autre. Cette augmentation reflète probablement la difficulté accrue de la recherche des évènements visuels lorsque la vidéo est dégradée (besoin de balayer plus largement la scène pour repérer les contacts entre les acteurs par exemple).

Des variations de la fréquence de clignement en fonction de l'interaction entre séquences, présence de dégradations et type de tâche ont également été constatées. Globalement, EBfreq tendait à diminuer en présence de dégradations sauf pour les contenus Opéra et Théâtre qui montraient le pattern inverse. La tendance à diminuer en présence de dégradations pourrait refléter un effort mental. La tendance à augmenter en présence de dégradations pourrait refléter un état de fatigue. Plus spécifiquement, une augmentation significative du nombre de clignements lors de la réalisation d'une des deux tâches (T1) de la séquence *Théâtre* a été observée. Cette augmentation pourrait refléter un état de fatigue lié à la réalisation de la tâche de comptage d'évènements auditifs en présence de la dégradation audio et avec un format 3D dégradant la qualité vidéo. Ainsi, la perte d'intelligibilité, d'autant plus gênante pour la réalisation d'une tâche de comptage de mots cibles, alliée à une perte de qualité vidéo pourrait avoir entraîné un effort mental suffisant pour engendrer un phénomène de fatigue.

Enfin, la réalisation de la tâche 2 de la séquence *Théâtre* a entraîné une augmentation de l'AED pouvant traduire une activation physiologique. Celle-ci serait liée à l'effort mental induit par la tâche de comptage d'évènements auditifs lors d'un contenu présentant des dégradations audio. Comme précédemment précisé, *Théâtre* présentait une qualité vidéo dégradée par la présentation 3D ainsi qu'une probable perte d'intelligibilité en raison du contenu fortement verbal (et diégétique) de la séquence. Il est intéressant de noter que cette séquence a influencé à la fois les indicateurs EBfreq (influence de T1 sur fatigue) et AED (influence de T2 sur effort). Il semblerait que la réalisation d'une tâche en présence de dégradations à la fois sur le signal audio (perte d'intelligibilité) et vidéo (+3D) entraîne une activation du système nerveux sympathique (dépenses énergétiques accrues). Ce résultat permet de confirmer H2.

J. CONCLUSIONS

Les résultats ont mis en avant une **influence de la qualité sur les performances**, celles-ci diminuant en présence de dégradations. Le questionnaire subjectif (qualité, émotions, QoE) proposé a également permis de différencier les séquences dégradées des séquences non dégradées. Globalement, la présence de dégradations a influencé **le niveau de qualité perçue (diminution), l'état émotionnel du spectateur (plus largement négatif) et les évaluations des dimensions de QoE** (augmentation de l'effort mental ressenti et diminution de l'intérêt, de la compréhension et de la qualité hédonique de la séquence qui était alors jugée comme moins stimulante, originale et agréable). Par ailleurs, l'évaluation de qualité (notes MOS) a indiqué que **la qualité vidéo a plus fortement influencé la qualité audiovisuelle que la qualité audio**. L'évaluation de la qualité vidéo était indépendante de la qualité audio en revanche, cette dernière a été influencée par la qualité vidéo. Enfin, **le type de séquences et la présence de dégradations ont influencé l'activité physiologique et oculaire** des spectateurs. Ce constat permet de croire que le protocole proposé est une piste prometteuse pour l'étude de l'influence de la qualité audiovisuelle à travers ce type de mesure. Plus généralement, un protocole de « double-tâche » (visualisation/tâche explicite) pourrait être l'approche appropriée pour étudier l'influence de la qualité audiovisuelle restituée, notamment en matière de coût pour le spectateur, et reportée à l'aide de mesures de performances, de mesures subjectives et des mesures psychophysiologiques.